ABSTRACT

CAI, YUHENG. Deep Learning-Based Automated Cell Detection and Classification in Fluorescence Microscopy Images. (Under the direction of Dr. Alon Greenbaum).

To comprehensively study tissues at the cellular level, both the quantity and spatial distribution of cell populations are critical as their functional roles are influenced by local environment and external cues. While immunohistology combined with advanced microscopy enables precise tissue labeling, manual cell counting becomes cumbersome with large numbers of labeled cells, necessitating automation of cell detection and classification for quantitative analysis. This dissertation innovatively utilized deep neural networks for object detection to accurately locate and classify labeled cells in fluorescence images, with simplified data annotation and novel data augmentation methods. These approaches enabled automated detection and localization of cells marked with whole-cell or cytoplasmic staining, facilitating precise quantitative analysis in 3D profiling of brains and cochleae.

The study initially employed RetinaNet, an object detection model, to locate and classify genetically labeled neurons and glia in Mosaic Analysis with Double Markers (MADM) mouse brain sections imaged by fluorescence microscopy, achieving an impressive average precision of 0.90. Subsequently, a deep learning-based pipeline built upon RetinaNet was developed to automatically detect neurons and astrocytes in entire mouse brains. Leveraging tissue clearing and light sheet fluorescence microscopy (LSFM), terabytes of data containing millions of labeled cells were acquired, making manual analysis impractical. To address this, the automated workflow was implemented to map labeled neurons and astrocytes in genetically distinct MADM mouse forebrains and register them to the Allen Brain Atlas (ABA) for region-wise analysis. Hierarchical analysis of region-wise statistics based on ABA annotation allowed to average the cell densities across larger brain areas that are morphologically and functionally connected.

Quantitative analysis revealed regional variations in astrocyte populations corresponding to genotypes given MADM-based deletion of epidermal growth factor receptor, consistent with previous studies relying on laborious manual counting and registrations, affirming the efficacy of the proposed workflow.

The research further extended to quantifying spiral ganglion neurons (SGNs) in gerbil and porcine cochleae. SGNs, crucial for transmitting auditory information to the brain, are situated deeper within the inner ear, specifically within the modiolus. Despite their clinical significance in understanding auditory function and potential therapeutic targets for hearing loss, SGNs have remained relatively understudied due to the technical challenges of imaging them within the bone. Previous methodologies using tissue clearing and LSFM, primarily focused on hair cells and supporting cells, leaving a notable gap in the comprehensive 3D imaging and quantification of SGNs, particularly within large animal models such as pigs. In response to this gap, a workflow was developed for immunostaining, imaging, and automated detection of all Type I SGNs in intact gerbil and porcine cochleae. Particularly, we introduced collagenase treatment to improve antibody penetration within the modiolus and achieved uniform staining quality. To enhance the accuracy of RetinaNet detection, the physical model of the imaging system was used to simulate defocus aberration that dominates LSFM. Our proposed workflow yielded cell counts comparable to manual analysis in gerbil datasets and provided valuable insights into Type I SGN counts in porcine cochlea. This comprehensive approach represents a significant advancement in the field, offering a robust methodology for studying and quantifying SGNs, thus enhancing our understanding of auditory neural circuitry, and paving the way for potential clinical applications.

In summary, the proposed deep learning-based cell detection workflows offer a versatile solution for spatial analysis of cellular populations, applicable not only to neuroscience studies but also to any tissue preparation containing labeled cell populations. These approaches hold promise for diverse fluorescence image datasets, providing researchers with valuable tools for comprehensive image analysis.

Deep Learning-Based Automated Cell Detection and Classification in Fluorescence Microscopy Images

by
Yuheng Cai

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Biomedical Engineering

Raleigh, North Carolina
2024

APPROVED BY:

Dr. Alon Greenbaum Committee Chair	Dr. Ke Cheng
Dr. Donald O. Freytes	Dr. H. Troy Ghashghaei
Dr. Shahar Z. Kovalsky	

DEDICATION

To my hamsters (maju, Huhu, bealt, Deggtart) and betta (Drow, Putao and Chuyi).

献给我亲爱的家人。

BIOGRAPHY

Yuheng (Rene) Cai was born and raised in Guangshui, Hubei, China. She finished her bachelor's degree in Biomedical Engineering at Shanghai Jiao Tong University in 2017. Then, she went to the UK and obtained a master's degree in Biomedical Research, specializing in Data Science, from Imperial College London in 2018. Her doctoral journey in Biomedical Engineering, supervised by Dr. Alon Greenbaum, began in August 2019 in the US.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my advisor, Dr. Alon Greenbaum, whose expertise, understanding, and patience added considerably to my graduate experience. Your guidance helped me navigate through the challenges of research and study, and I am immensely thankful for your continuous support.

I also wish to extend my thanks to the members of my advisory committee, Dr. Ke Cheng, Dr. Donald O. Freytes, Dr. H. Troy Ghashghaei, and Dr. Shahar Z. Kovalsky, for their insightful feedback and valuable suggestions throughout this process.

My sincere appreciation goes to all the members of Greenbaum Lab, especially Dr. Adele Moatti, Dr. Mani Ratnam Rai, and Chen Li, for the stimulating discussions, for the time we were working together, and for all the fun we had. I am thankful for the opportunity to work alongside such dedicated and talented individuals.

A special thanks to our collaborators, Dr. Douglas C. Fitzpatrick, Dr. Kendall A. Hutson, Dr. Tobias Kaeser, Dr. Andrew J. Newell, and their lab members for their help and contributions to our joint efforts.

I would also like to express my gratitude to my undergraduate and master’s advisors and mentors, Dr. Liming Li, Dr. Qian Wang, Dr. Puming Zhang; Dr. Timothy M. D. Ebbels, Dr. Elizabeth J. Want, Dr. Gonçalo Graça, for their guidance and mentorship throughout my academic journey.

I cannot express enough thanks to my family and friends, my significant other Dr. Ziyang Xie, and my hamster and betta friends :), for their love and support throughout this journey. Your unwavering support has been my anchor, and I am profoundly grateful for everything you have done for me.

I would also like to thank the laboratory animals for their invaluable role in advancing scientific knowledge.

This dissertation stands as a milestone in my academic journey, and I am thankful for everyone who played a role in reaching this point. Your support and encouragement were pivotal in my pursuit of this achievement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

Appendix A: CHAPTER 2

Appendix B: CHAPTER 3

Appendix B: CHAPTER 3

CHAPTER 1: Introduction

1. Background

1.1 Tissue clearing of biological specimens

Fluorescence microscopy is a type of microscopy using fluorescence to visualize biological structures and molecules, showing advantages in specificity, multiplexity, and sensitivity [1]. With fluorescence microscopy, researchers can study cells and tissues at subcellular resolution with specific staining of targets of interest using fluorescent dyes and fluorescently labeled antibodies, allowing for visualization of individual cells and their components, providing g insights into their structure and function. Yet, challenges arise in fluorescence imaging of thick tissue samples, as light gets absorbed, scattered, and attenuated when passing through biological tissues, resulting in reduced signal intensity and poorer image quality in deep tissue regions. To overcome this limitation, tissues are cut into sections during sample preparation, thus reducing optical aberrations and scattering, and allowing for uniform penetration of fluorescent dyes and antibodies. However, using this sectioning approach, 3D structural information of the tissue is lost. Tissue sectioning may also introduce artifacts such as tissue distortion, tearing, or folding, as well as damage to the tissue, making it challenging to reconstruct the original 3D structure accurately. Such limitations necessitate the development of specialized sample preparation techniques, specifically tissue clearing, that enable deep imaging of whole-mount tissues.

By rendering biological tissue almost transparent, tissue clearing allows researchers to visualize the internal structures of tissues with high clarity and resolution. Tissue clearing involves treating biological tissues with chemical reagents or solutions to remove or minimize light scattering and refractive index mismatches, thus improving light penetration and enhancing imaging depth. Figure 1 shows examples of tissue cleared soft tissue (mouse brain) and hard tissue (pig cochlea).

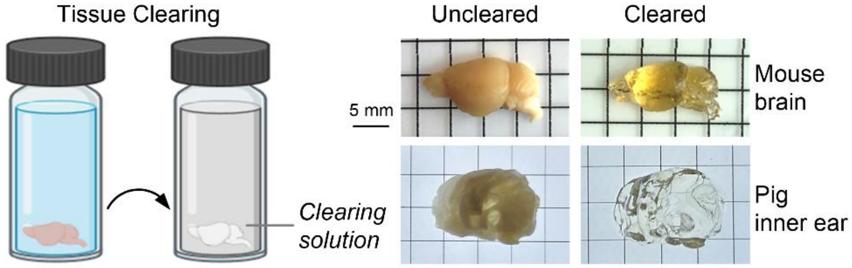


Figure 1. Tissue clearing renders entire biological samples transparent. Tissue clearing is achieved by matching the refractive index of the sample with clearing reagents. A mouse brain and a pig cochlea before and after tissue clearing are shown. The samples were imaged in the air.

Various tissue clearing methods are available including organic solvent-based, simple immersion, hyperhydration, and hydrogel embedding [2]. Solvent-based methods remove water and lipids from the sample and immerse the sample in high refractive index organic solvents like benzyl alcohol, benzyl benzoate, or dibenzyl ether, providing good clarity but often quenching the emission of fluorescent dyes or proteins which limits their utility. To preserve fluorescent labeling in the sample, aqueous-based clearing is pursued by immersion of tissue in refractive index matched solutions. Clearing can be achieved by simple immersion for small samples that still contain water molecules and lipids, which may take days to months and results in slightly worse clearing quality. An alternative approach is to remove the lipids by detergents and reduce the refractive index of tissue samples by hyperhydration with the usage of urea or formamide, thus produces better clearing for thick tissue samples. Yet, the extensive used of detergents in such hyperhydration methods can result in protein loss in a high degree, which lowers the concentrations of epitopes and potentially weakens immunostaining. Hydrogel embedding, performed before lipid removal and clearing, preserves tissue structure and supports fluorescence labeling and immunostaining.

Tissue clearing enables researchers to study the 3D architecture, cellular organization, and molecular distribution within tissues, providing valuable insights into biological processes, disease mechanisms, and therapeutic intervention.

1.2 3D Imaging of tissue cleared samples

To image tissue cleared samples, confocal fluorescence microscopy (CFM) and light sheet fluorescence microscopy (LSFM) are commonly used, which are able to exclude out-of-focus fluorescence signals above and below the focal plane, thus increasing image sharpness and clarity when imaging thick biological samples. As shown in Figure 2, CFM illuminates the sample with a focused laser beam to excite the fluorophores within the specimen and uses a pinhole aperture to exclude emitted light that is out of the focal plane, requiring point-by-point laser scanning during image acquisition. LSFM utilizes a thin layer of laser light, so called a light sheet, to selectively illuminate a plane within the sample, while fluorescence emitted from the illuminated plane is captured by a detection objective positioned orthogonal to the light sheet. Note that during imaging, the illuminated plane should be overlapped with the focal plane of the detection objective to obtain sharp images.

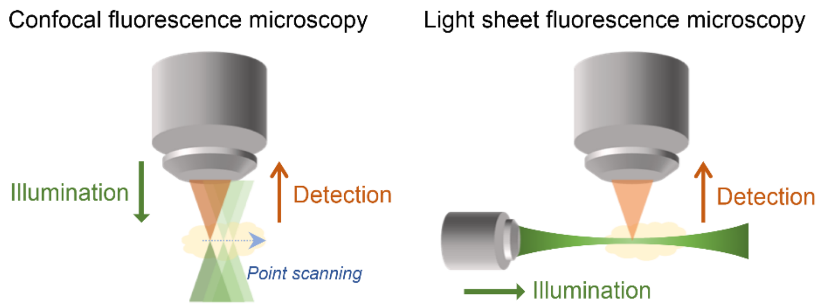


Figure 2. Schematic overview of confocal fluorescence microscopy and light sheet fluorescence microscopy.

Compared to CFM, LSFM shows advantages including reduced phototoxicity, minimized photobleaching, and high imaging speed, thus making it well-suited for imaging of larger volumes of tissue cleared samples which may take hours to days [3–6]. Leveraging tissue clearing and LSFM, researchers have studied neuronal activity in whole mouse brain [7], neuronal projections and skull-meninge connections in whole mouse body [8], cellular architecture within entire human organs [6], and more [5]. Whie maintaining all the structural information, 3D imaging of tissue cleared samples can generate terabytes of images to process and analyze.

1.3 Deep learning for image analysis

For either 2D or 3D imaging, challenges arise in quantitative analysis of the acquired datasets, e.g., 3D imaging can generate terabytes of data and millions of labeled cells to analyze, thus making manual analysis time-consuming, laborious and error prone. To accelerate and automate image analysis, machine learning algorithms, particularly recently developed deep learning models can be used, which have shown their efficiency and efficacy in various image processing tasks including classification, object detection, segmentation and more [9].

Researchers have been exploring artificial intelligence for decades, aiming to create systems or machines to mimic human intelligence in complex tasks like perception, understanding natural language, learning and more. Machine learning is a subset of artificial intelligence that involves the development of algorithms and models that enable computers to learn from data. Deep learning is a type of machine learning that employs artificial neural networks with multiple layers of interconnected nodes (neurons) organized in a hierarchical fashion, allowing the network to learn complex representations of data through successive layers of abstraction.

Compared to other machine learning methods, deep learning is able to automatically learn complicated feature patterns from data and achieves state-of-the-art performance on complex tasks such as image processing, speech recognition, natural language processing, and more. Specifically, deep learning has revolutionized image analysis [10], with key applications in image classification to categorize images into predefined classes, object detection to identify and locate objects of interest within images, semantic segmentation to assign pixel-wise labels to delineate different objects or regions of interest, image synthesis to generate images that are indistinguishable from real data, etc. Recently, deep learning models have been applied to biomedical research, including classification of lung carcinoma [11], detection of blood cells [12], segmentation of nucleus with style transfer capability [13], investigation of cancer metastasis in mice [14], and so on.

1.4 Mosaic Analysis with Double Markers

Mosaic Analysis with Double Markers (MADM) is a sophisticated genetic technique used in developmental biology to study cell lineage and fate determination in vivo [15]. MADM enables researchers to induce somatic recombination events within tissues or organisms, generating mosaic patterns of genetically modified cells with distinct genetic markers or mutations. One of the key features of MADM is the use of two different fluorescent protein markers, green fluorescent protein (GFP) and red fluorescent protein (RFP), to label cells with distinct genetic alterations. This enables simultaneous visualization and tracking of mutant and wild-type cells within the same tissue, providing valuable insights into cell lineage relationships and clonal dynamics. Figure 3 shows the sparse labeling of progenitors in mice with MADM enabled epidermal growth factor receptor (Egfr) deletion [16]. Cre-mediated interchromosomal mitotic recombination events lead to definite labeling of cells at low densities, exhibiting two distinct segregation patterns. Cells expressing GFP (green) and tdTomato (red, an RFP) originate from one type of division, while yellow and non-fluorescent siblings arise from another type of division. All the cells in the control mice (+/+) are wild type (*WT*). In the *F/+* mice where a single floxed allele is introduced, green cells lack Egfr (*Egfr-null*) and red cells remain *WT*.

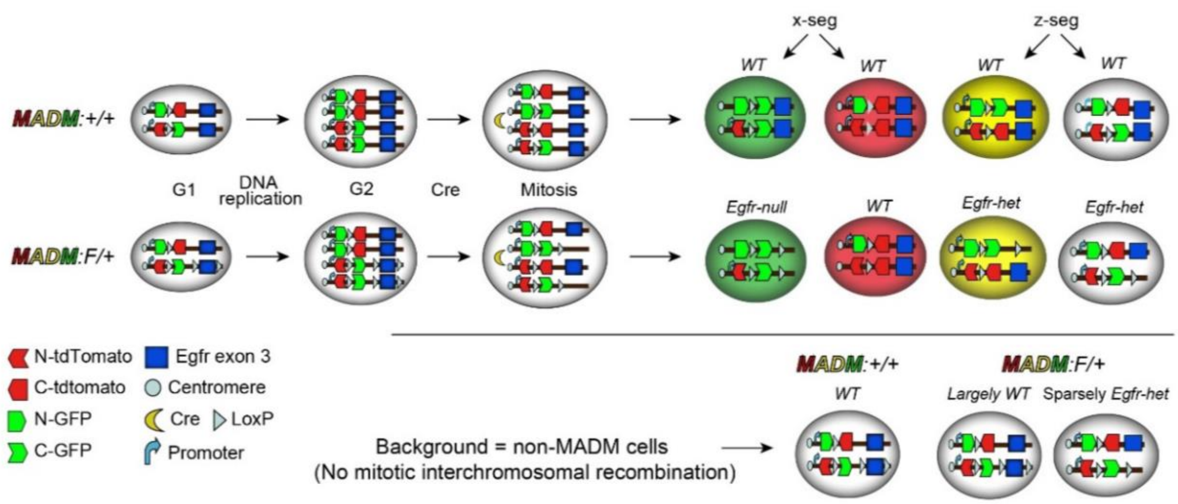


Figure 3. Molecular basis of labeling progenitors with two fluorescent proteins using Mosaic Analysis with Double Markers (MADM), with deletion of epidermal growth factor receptor (Egfr). Adapted from [16]. Cre-mediated interchromosomal mitotic recombination result in two segregation patterns (x-seg and z-seg), generating clonal siblings in different colors: green (GFP), red (tdTomato) and yellow (both reporters expressed). In control MADM mice (+/+), all cells will be wild type (*WT*). Introducing a single floxed allele (*F/+*) leads to green cells lacking Egfr (*Egfr-null*), while red cells remain *WT*. GFP, green fluorescent protein. tdTomato, a red fluorescent protein.

2 Motivation and problem statement

The special distribution of cells within tissues provides valuable insights into tissue organization, architecture, and function, advancing our understanding of tissue biology and disease pathogenesis [17,18]. Therefore, it is important to identify cell populations and simultaneously find their locations within the tissue. To perform such an analysis, the utilization of object detection networks shows great promise as they not only identify objects of interest but also provide locations of the objects within the image, after which object counts can be easily obtained [19]. It is worth noting that compared to other deep learning models, object detection networks have not been extensively utilized for analyzing biomedical images. Among state-of-the-art object detection models, RetinaNet [20] shows superior performance on benchmark real world datasets [19] and fluorescence microscopy images [21]. Therefore, in this dissertation, RetinaNet is selected as the object detection model and applied to fluorescence microscopy images obtained from mouse brain sections, tissue cleared mouse brains, gerbil cochleae, and porcine cochlea. With the long-term goal of building an automated and robust quantitative analysis platform for fluorescence microscopy images, in this research we develop deep learning-based workflows to automatically process and analyze fluorescence microscopy images. Specifically, we design and apply automated pipelines to quantitatively analyze images acquired from different specimens with distinct sample preparations:

MADM mouse brain sections. MADM models facilitate simultaneous labeling and genetic manipulation in somatic cell clones derived during development. This approach offers the advantage of labeling neuron and glia genotypes in tissue through the expression of two fluorescent proteins. MADM is widely employed in developmental research focusing on genetic factors affecting neurogenesis [22–24] and gliogenesis [25–27]. However, analyzing the large number of labeled cells in tissue sections poses challenges. Therefore, an automated method is required to detect and classify MADM-labeled cells (whole-cell staining) in fluorescence images of mouse brain sections, offering adaptability to various studies involving fluorescence images containing labeled cell populations.

Entire tissue cleared MADM mouse brains. By combining tissue clearing with LSFM, we can label specific targets within intact organs and conduct quantitative and structural analyses by correlating cells with structural references. While previous research has established pipelines

[7,28–31] for automated analysis of tissue cleared mouse brains, these primarily focus on single color channels and/or nuclear fluorescence signals. Thus, there is a necessity for a novel approach to address cases where whole cells are stained, colocalization of color channels is crucial, and labeled cells can be classified based on morphology. We will explain in Chapter 3 why color localization is such a challenge in tissue cleared samples that are typically imaged using LSFM. Such a workflow is adaptable for various neuroscience studies requiring quantification of elements in the mouse brain using tissue clearing and immunofluorescence labeling.

Intact tissue cleared gerbil and porcine cochleae. Spiral ganglion neurons (SGNs), responsible for transmitting signals from the inner ear to the brain, have been relatively understudied due to the technical challenges of imaging them in 3D as they are located within the cochlea's modiolus. Investigating SGNs can yield crucial insights into auditory functions and aid in the identification of therapeutic targets for hearing loss that affects nearly one-fifth of the global population. Consequently, there is a pressing need for a framework to immunostain, image, and analyze the SGN population throughout the entire cochlea. Such a framework can facilitate the characterization of treatment effects at a cellular level and be adaptable to various studies related to the inner ear.

3 Dissertation outline

The remainder of the dissertation is organized as follows. Chapter 2 describes the utilization of an object detection model, specifically RetinaNet, to detect neurons and glia in MADM mouse brain sections imaged with a slide scanner and a CFM. Unique data augmentation methods such as color swap and saturation are employed to improve class balance and model performance. In Chapter 3, an automated workflow is developed to immunostain, image, and detect neurons and astrocytes in entire tissue cleared MADM mouse brains, with the ability to register the detected cells to Allen Brain Atlas for structural analysis. Regional effects of conditional deletion of Egfr on astrocyte population are investigated, showing trends consistent with manual analysis. In Chapter 4, a pipeline to immunostain, image, and analyze Type I SGNs in gerbil and porcine cochlea is developed. The pipeline yields cell counts comparable to manual analysis and report numbers of Type I SGNs in the porcine cochlea. Chapter 5 summarizes the current findings and discusses the future work.

REFERENCES

1. Sanderson MJ, Smith I, Parker I, Bootman MD. Fluorescence Microscopy. Cold Spring Harb Protoc. 2014 Oct 1;2014(10):pdb.top071795.
2. Richardson DS, Lichtman JW. Clarifying Tissue Clearing. Cell. 2015 Jul;162(2):246–57.
3. Kubota SI, Takahashi K, Nishida J, Morishita Y, Ehata S, Tainaka K, et al. Whole-Body Profiling of Cancer Metastasis with Single-Cell Resolution. Cell Reports. 2017 Jul 5;20(1):236–50.
4. Moatti A, Li C, Sivadanam S, Cai Y, Ranta J, Piedrahita JA, et al. Ontogeny of cellular organization and LGR5 expression in porcine cochlea revealed using tissue clearing and 3D imaging. iScience. 2022 Aug 19;25(8):104695.
5. Ueda HR, Ertürk A, Chung K, Gradinaru V, Chédotal A, Tomancak P, et al. Tissue clearing and its applications in neuroscience. Nat Rev Neurosci. 2020 Feb;21(2):61–79.
6. Zhao S, Todorov MI, Cai R, -Maskari RA, Steinke H, Kemter E, et al. Cellular and Molecular Probing of Intact Human Organs. Cell. 2020 Feb 20;180(4):796-812.e19.
7. Renier N, Adams EL, Kirst C, Wu Z, Azevedo R, Kohl J, et al. Mapping of Brain Activity by Automated Volume Analysis of Immediate Early Genes. Cell. 2016 Jun 16;165(7):1789–802.
8. Cai R, Pan C, Ghasemigharagoz A, Todorov MI, Förstera B, Zhao S, et al. Panoptic imaging of transparent mice reveals whole-body neuronal projections and skull–meninges connections. Nat Neurosci. 2019 Feb;22(2):317–27.
9. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015 May;521(7553):436–44.
10. Voulodimos A, Doulamis N, Doulamis A, Protopapadakis E. Deep Learning for Computer Vision: A Brief Review. Comput Intell Neurosci. 2018;2018:7068349.

11. Kanavati F, Toyokawa G, Momosaki S, Rambeau M, Kozuma Y, Shoji F, et al. Weakly-supervised learning for lung carcinoma classification using deep learning. Sci Rep. 2020 Jun 9;10(1):9297.

12. Jiang Z, Liu X, Yan Z, Gu W, Jiang J. Improved detection performance in blood cell count by an attention-guided deep learning method. OSA Continuum. 2021 Feb 15;4(2):323.

13. Hollandi R, Szkalisity A, Toth T, Tasnadi E, Molnar C, Mathe B, et al. nucleAIzer: A Parameter-free Deep Learning Framework for Nucleus Segmentation Using Image Style Transfer. Cell Systems. 2020 May 20;10(5):453-458.e6.

14. Pan C, Schoppe O, Parra-Damas A, Cai R, Todorov MI, Gondi G, et al. Deep Learning Reveals Cancer Metastasis and Therapeutic Antibody Targeting in the Entire Body. Cell. 2019 Dec;179(7):1661-1676.e19.

15. Zong H, Espinosa JS, Su HH, Muzumdar MD, Luo L. Mosaic Analysis with Double Markers in Mice. Cell. 2005 May;121(3):479–92.

16. Zhang X, Mennicke CV, Xiao G, Beattie R, Haider MA, Hippenmeyer S, et al. Clonal Analysis of Gliogenesis in the Cerebral Cortex Reveals Stochastic Expansion of Glia and Cell Autonomous Responses to Egfr Dosage. Cells. 2020 Dec 11;9(12).

17. Ruprecht V, Monzo P, Ravasio A, Yue Z, Makhija E, Strale PO, et al. How cells respond to environmental cues – insights from bio-functionalized substrates. J Cell Sci. 2017 Jan 1;130(1):51–61.

18. Suárez LE, Markello RD, Betzel RF, Misic B. Linking Structure and Function in Macroscale Brain Networks. Trends in Cognitive Sciences. 2020 Apr 1;24(4):302–15.

19. Zou Z, Shi Z, Guo Y, Ye J. Object Detection in 20 Years: A Survey. arXiv:190505055 [cs] [Internet]. 2019 May 15 [cited 2021 Feb 28]; Available from: http://arxiv.org/abs/1905.05055

20. Lin TY, Goyal P, Girshick R, He K, Dollár P. Focal Loss for Dense Object Detection. arXiv:170802002 [cs] [Internet]. 2018 Feb 7 [cited 2019 Dec 9]; Available from: http://arxiv.org/abs/1708.02002

21. Waithe D, Brown JM, Reglinski K, Diez-Sevilla I, Roberts D, Eggeling C. Object detection networks and augmented reality for cellular detection in fluorescence microscopy. Journal of Cell Biology. 2020 Oct 5;219(10):e201903166.

22. Gao P, Postiglione MP, Krieger TG, Hernandez L, Wang C, Han Z, et al. Deterministic progenitor behavior and unitary production of neurons in the neocortex. Cell. 2014 Nov 6;159(4):775–88.

23. Johnson CA, Ghashghaei HT. Sp2 regulates late neurogenic but not early expansive divisions of neural stem cells underlying population growth in the mouse cortex. Development. 2020 Feb 21;147(4).

24. Laukoter S, Beattie R, Pauler FM, Amberg N, Nakayama KI, Hippenmeyer S. Imprinted Cdkn1c genomic locus cell-autonomously promotes cell survival in cerebral cortex development. Nat Commun. 2020 Jan 10;11(1):195.

25. Beattie R, Postiglione MP, Burnett LE, Laukoter S, Streicher C, Pauler FM, et al. Mosaic Analysis with Double Markers Reveals Distinct Sequential Functions of Lgl1 in Neural Stem Cells. Neuron. 2017 May 3;94(3):517-533.e3.

26. Zhang X, Mennicke CV, Xiao G, Beattie R, Haider MA, Hippenmeyer S, et al. Clonal Analysis of Gliogenesis in the Cerebral Cortex Reveals Stochastic Expansion of Glia and Cell Autonomous Responses to Egfr Dosage. Cells. 2020 Dec 11;9(12).

27. Zhang X, Xiao G, Johnson C, Cai Y, Horowitz ZK, Mennicke C, et al. Bulk and mosaic deletions of Egfr reveal regionally defined gliogenesis in the developing mouse forebrain. iScience. 2023 Mar 17;26(3):106242.

28. Frasconi P, Silvestri L, Soda P, Cortini R, Pavone FS, Iannello G. Large-scale automated identification of mouse brain cells in confocal light sheet microscopy images. Bioinformatics. 2014 Sep 1;30(17):i587–93.

29. Silvestri L, Müllenbroich MC, Costantini I, Di Giovanna AP, Mazzamuto G, Franceschini A, et al. Universal autofocus for quantitative volumetric microscopy of whole mouse brains. Nat Methods. 2021 Aug;18(8):953–8.

30. Matsumoto K, Mitani TT, Horiguchi SA, Kaneshiro J, Murakami TC, Mano T, et al. Advanced CUBIC tissue clearing for whole-organ cell profiling. Nat Protoc. 2019 Dec;14(12):3506–37.

31. Krupa O, Fragola G, Hadden-Ford E, Mory JT, Liu T, Humphrey Z, et al. NuMorph: Tools for cortical cellular phenotyping in tissue-cleared whole-brain images. Cell Reports. 2021 Oct 12;37(2):109802.

CHAPTER 2: Detection and classification of neurons and glial cells in the MADM mouse brain using RetinaNet

Yuheng Cai[1,2], Xuying Zhang[2,3], Shahar Z. Kovalsky[4], H. Troy Ghashghaei[2,3], Alon Greenbaum[1,2,5*]

[1] Joint Department of Biomedical Engineering, North Carolina State University and University of North Carolina at Chapel Hill, Raleigh, North Carolina/Chapel Hill, United States of America

[2] Comparative Medicine Institute, North Carolina State University, Raleigh, North Carolina, United States of America

[3] Department of Molecular Biomedical Sciences, North Carolina State University, Raleigh, North Carolina, United States of America

[4] Department of Mathematics, University of North Carolina, Chapel Hill, North Carolina, United States of America

[5] Bioinformatics Research Center, North Carolina State University, Raleigh, North Carolina, United States of America

* Corresponding author

E-mail: greenbaum@ncsu.edu; greenbaum@unc.edu (AG)

Cai, Y., Zhang, X., Kovalsky, S. Z., Ghashghaei, H. T., & Greenbaum, A. (2021). Detection and classification of neurons and glial cells in the MADM mouse brain using RetinaNet. *Plos one*, *16*(9), e0257426.

https://doi.org/10.1371/journal.pone.0257426

1 Abstract

The ability to automatically detect and classify populations of cells in tissue sections is paramount in a wide variety of applications ranging from developmental biology to pathology. Although deep learning algorithms are widely applied to microscopy data, they typically focus on segmentation which requires extensive training and labor-intensive annotation. Here, we utilized object detection networks (neural networks) to detect and classify targets in complex microscopy images, while simplifying data annotation. To this end, we used a RetinaNet model to classify genetically labeled neurons and glia in the brains of Mosaic Analysis with Double Markers (MADM) mice. Our initial RetinaNet-based model achieved an average precision of 0.90 across six classes of cells differentiated by MADM reporter expression and their phenotype (neuron or glia). However, we found that a single RetinaNet model often failed when encountering dense and saturated glia clusters, which show high variability in their shape and fluorophore densities compared to neurons. To overcome this, we introduced a second RetinaNet model dedicated to the detection of glia clusters. Merging the predictions of the two computational models significantly improved the automated cell counting of glia clusters. The proposed cell detection workflow will be instrumental in quantitative analysis of the spatial organization of cellular populations, which is applicable not only to preparations in neuroscience studies, but also to any tissue preparation containing labeled populations of cells.

2 Introduction

The functional role of a cell is highly dependent on its gene expression, local environment, and external cues [1,2]. The two latter factors are directly related to the spatial location of the cell (e.g., layers in the neocortex, regions in the hippocampus and more). Therefore, in a tissue section, both the number of labeled cells and their spatial distribution are of paramount importance [3–5]. To build these spatial distribution maps, two technologies have been instrumental: (*i*) Sophisticated and automated microscopes (e.g., slide scanners) that facilitate high-throughput data acquisition of biomedical specimens. (*ii*) Tissue labeling methods that include immunohistochemistry, *in situ* hybridization, transgenic reporter mice and more.

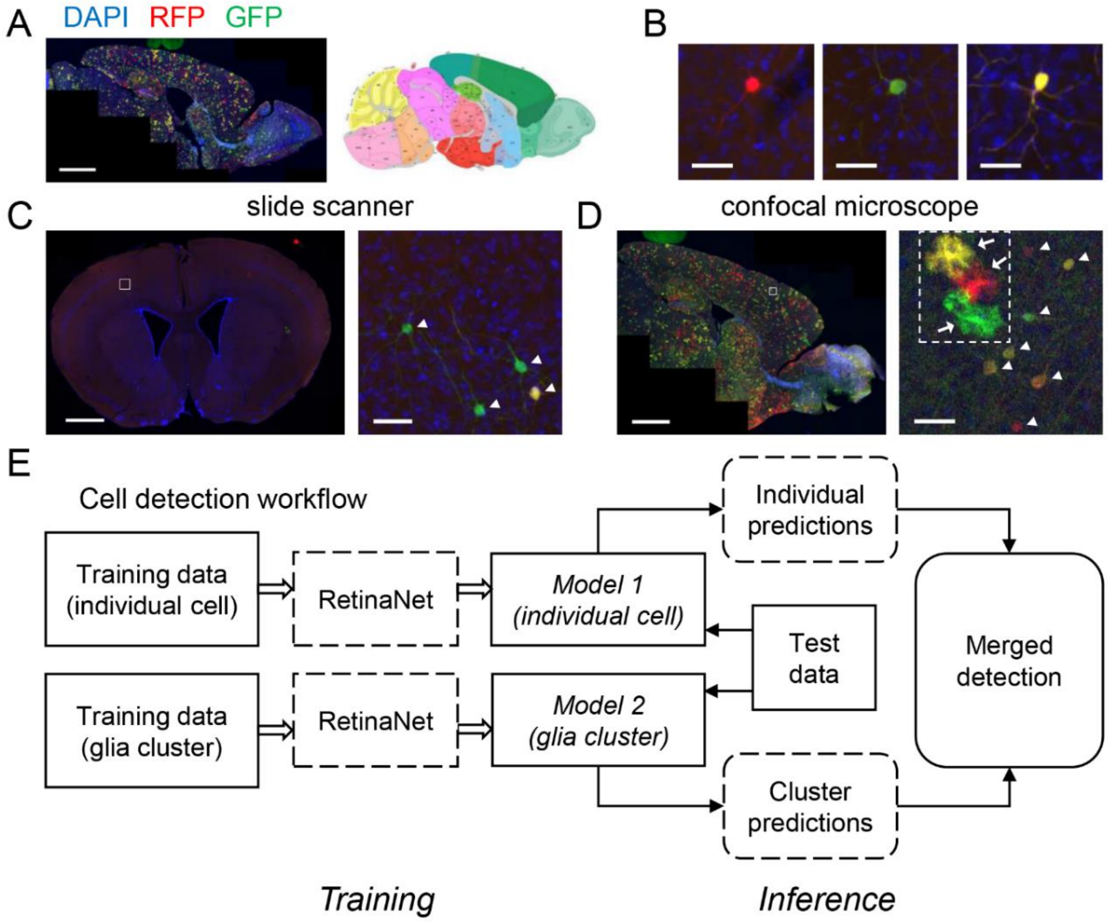
Here our goal is to detect and classify cells in images of brain sections obtained from Mosaic Analysis with Double Markers (MADM) mice. MADM allows for simultaneous labeling and genetic manipulation in developmentally derived clones of somatic cells [6]. We and others have extensively used MADM alleles in developmental studies on the roles of various genetic factors, which are involved with neurogenesis [7–9] and gliogenesis [10,11]. An advantage of MADM is that neurons and glia with distinct genotypes are permanently labeled by expression of two fluorescent proteins. Furthermore, MADM labeling occurs in sparse populations such that the entire morphology of individual cells can be easily resolved using microscopy (Figure 1A and B). However, in some MADM preparations an entire brain section can contain large numbers of cells despite the sparsity of genetic labeling relative to the total number of cells, which can render manual cell counting tedious and error prone. Hence, the automation of cell detection and classification is vital to boost throughput and unbiased approaches necessary for quantification of complex tissues such as MADM brain sections.

To address this gap, machine-learning algorithms have been utilized to automate and accelerate data processing. Among them, deep learning has outperformed many conventional machine-learning algorithms in multiple tasks e.g., natural language processing, computer vision, speech recognition and more [12]. In computer vision, various deep learning architectures have been employed to address different tasks: For example, convolutional neural networks (CNNs) are used for image classification, Region based CNNs (R-CNNs) are used for object detection, and fully convolutional networks (FCN) are used for semantic segmentation [13]. These deep neural networks (DNNs) have also been applied to biomedical data and showed great promise in, for example, classification of breast cancer histopathology images (CNN) [14], segmentation of whole mouse brain vasculature (3D CNN) [15], and detection of blood cells (You Only Look Once network) [16]. Compared to other network architectures, object detection networks have yet to be extensively used for analysis of biological data. This is surprising as the training of these networks is relatively simple, and their ability to localize and classify targets even in real-time is excellent [17,18]. RetinaNet [19], You Only Look Once (YOLO) [20], Single Shot Detector (SSD) [21], Faster R-CNN [22] are the state-of-art object detection models developed in recent years [18]. Without limitation on the inference time, which can be up to 200 milliseconds, RetinaNet shows superior or comparable results on benchmark datasets [23]. For detection of cells in fluorescence microscopy, recent report have shown that both YOLOv2 and RetinaNet perform well [24]. In our implementation, RetinaNet is selected for cell detection given its slightly better performance on multiple benchmark datasets and its higher performance in dense object detection [19].

Here we present an automatic cell detection workflow using RetinaNet models to analyze brain sections that were obtained from MADM mice (Figure 1C-E). We put forth training considerations that accelerate the training process and improve model performance. To diversify the training data and generalize our results, images were acquired by either a confocal fluorescence microscope (CFM) or a slide scanner, all of which contained different cellular densities (i.e., genetically different MADM mice). Novel data augmentation methods were also used to compensate for the imbalance in MADM cell numbers due to genotypic differences in the training dataset. To resolve issues with detection and classification of high-density clusters of

MADM labeled glia, a unique workflow was designed which incorporated two RetinaNet models, resulting in superior performance compared to a single RetinaNet model trained to detect clusters.

Figure 1. Automated cell and cluster detection in MADM brains. (**A)** Confocal micrograph of a sagittal section from a month-old Nestin-cre, MADM-11 (MADM) mouse forebrain (left) and the corresponding annotated map from the Allen Brain Atlas (right; Image credit: Allen Institute). Scale bar, 1500 µm. GFP – green fluorescent protein, RFP – red fluorescent protein, DAPI – 4',6-diamidino-2-phenylindole. (**B)** Three isolated neurons captured in the MADM brain where green (enhanced GFP), red (tdTomato) and yellow (both reporters expressed) cells are derived from distinct clones of progenitors earlier during development [8,10,25,26]. Scale bars, 40 µm. (**C)** A representative coronal MADM section with only a single clone of cells labeled and later imaged by a slide scanner (left; scale bar, 1000 µm). Sections were obtained from MADM brains in which red, green, and yellow clones were labeled in the late-stage embryo at very low densities using a Nestin-creER transgene [11]. Boxed area demarcates the zoomed image on the right. Scale bar, 50 µm. (**D)** A representative confocal image of another MADM brain (left; scale bar, 1250 µm) and zoomed image to the right (right; scale bar, 50 µm). Two main types of cells can be seen in C and D: neurons and glia marked with arrowheads and arrows, respectively. The white dashed frame in D indicates a glia cluster. (**E)** The cell detection workflow. To localize and classify each cell, an object detection network (RetinaNet) was utilized. To detect dense and saturated glia clusters, two RetinaNet models were trained, one to detect individual cells with different colors, and the other to detect only glia clusters. In the inference stage, predictions of individual cells and glia clusters were merged to obtain final output.



A
DAPI RFP GFP
B
C
slide scanner
D
confocal microscope
E
Cell detection workflow
Training data (individual cell)
RetinaNet
Model 1 (individual cell)
Individual predictions
Test data
Merged detection
Training data (glia cluster)
RetinaNet
Model 2 (glia cluster)
Cluster predictions
Training
Inference

3 Materials and methods

3.1 Data generation

MADM-11 mice (The Jackson Laboratory, Bar Harbor, USA; #013751, #013749) were crossed to Nestin-cre mice (bred from MADM-11 mice) using the breeding scheme described previously [11] and harvested at the age of one month under the regulations and approval from the Institutional Animal Care and Use Committee at North Carolina State University. Mice were housed in a 12-h light:dark cycle with ad libitum access to food and water. MADM mice were deeply anesthetized by Avertin overdose (2,2,2 tribromoethanol; 7.5 mg/g body weight), perfused intracardially with 4% paraformaldehyde (PFA) in phosphate buffer saline (PBS, 0.1 M, hereafter), and brains were dissected and submerged in 4% PFA in PBS at 4 °C overnight. Brains were embedded in 3% low melting point DNA-grade agarose in PBS and serial 50 µm sections were collected using a vibratome (Leica VT1000S, Leica, Buffalo Grove, USA). Floating serial sections were washed with PBS and blocked for 1 h at room temperature in blocking buffer (10% normal donkey or goat serum, 1% Triton X-100, PBS). Sections were incubated with primary anti-GFP (Green Fluorescent Protein, Abcam, Cambridge, MA; ab13970, 1:2000) and Rabbit anti-RFP (Red Fluorescent Protein, Abcam, ab62341, 1:500) antibodies diluted in 0.1% blocking buffer overnight at 4 °C, followed by 3 5-min washes with PBS at room temperature the next day. Alexa Fluor goat anti rabbit Cy3 (Thermo Fisher Scientific, Waltham, USA; A10520, 1:1000), Alexa Fluor goat anti-chicken 488 (Thermo Fisher Scientific, A11039, 1:1000) secondary antibodies were diluted in blocking buffer and incubated with the serial sections for 1 h at room temperature, followed by 3 washes with PBS. Sections were counterstained with the DNA marker (4',6-diamidino-2-phenylindole; DAPI) at 1:2000 during the secondary incubation. Sections were mounted onto glass slides and coverslipped with

Faramount aqueous mounting medium (Dako, Agilent Technologies, Santa Clara, USA). Images of the MADM forebrains were acquired using an FV1000 confocal microscope (Olympus, Waltham, USA) or a slide scanner (VS120, Olympus, Waltham, USA).

3.2 Data annotation

The training data and test data were generated separately from 16 different mouse brain samples (52 brain sections), of which 9 mouse brains were imaged by a slide scanner and the rest were imaged by a confocal microscope. For training, we labeled 2009 individual cells (1219 neurons and 790 glia) and 168 glia clusters from 39 brain sections. For testing, we labeled 551 individual cells (346 neurons and 205 glia) and 48 glia clusters from 13 brain sections. Annotations of individual cells were generated using ilastik version 1.3.3 [27], ImageJ [28], and a customized graphical user interface (GUI) built in Python. The ilastik pixel classification workflow was used to distinguish cells from background for preprocessing. The interactive training process in ilastik allows users to monitor the output and adjust the labels until satisfactory results are obtained. A representative brain section image in the slide scanner dataset was used to train the algorithm in ilastik. After training, batch processing was done on all brain section images. The output probability maps from ilastik were imported into Python to extract centroids of each probable cell region. The centroids were then manually adjusted in ImageJ to precisely locate each cell, add centroids for miss detected cells, and remove false positives. The GUI was used to quickly add labels to each cell. Then the coordinates of fixed size bounding boxes were generated according to centroid locations. A Python script was used to export such annotations into formats compatible with the requirement of RetinaNet.

Glia clusters were annotated in LabelImg (https://github.com/tzutalin/labelImg), a labeling tool in Python. A Python script was used to transform the output format into RetinaNet-compatible format.

3.3 Data augmentation

Color swap was realized by swapping the RFP channel and the GFP channel of each image. During image acquisition, the output signal intensities are proportional to the input laser power. Therefore, to simulate the situation of saturation, the intensities in each channel were multiplied by a factor of 1.5, and a ceiling function was used to emulate saturation. The constant factor of 1.5 was empirically selected.

3.4 Training environments

To train the models NCSU Henry2 cluster was used, as well as UNC Longleaf cluster. For testing, a Lenovo ThinkStation P520 Workstation with one Quadro P1000 graphic processing unit (GPU) was used.

3.5 Object detection model

RetinaNet repository cloned from the source (https://github.com/fizyr/keras-retinanet) was modified for this work (https://github.com/yccc12/keras-retinanet). A pre-trained ResNet50 was used as the backbone. Zero-centering was used as a pre-processing step, as two types of microscopes were used to acquire the data, and zero-centering showed better results in comparison with normalization. Classical data augmentation strategies such as geometrical transforms and noise injection were applied. The initial learning rate was 0.0001 and the batch size was four. Using an Adam optimizer, all the models were trained for 50 epochs, where loss plateaus could be reached. To reduce the effect of randomness, which is inherent to the training process, the training process was repeated three times on the same training data and resulted in three independently trained networks. These networks were tested on the same test data, and their average precision (AP) results were averaged for comparison.

3.6 Area-based counting

For each image patch, the pixels were divided into two groups, pixels that were within an object bounding box and pixels that belonged to the background. For each channel, the mean and standard deviation (SD) of the background (BKG) pixels were calculated. The pixels that belonged to detected neurons within clusters were masked out. Then for each detected glia cluster, the standard deviation of pixels was calculated within each cluster. Afterwards, thresholding was used to detect pixels that belonged to cells within the cluster :

$$\boldsymbol{Threshold} = \frac{\mathbf{1}}{\mathbf{2}} \cdot (\boldsymbol{mean_{BKG}} + \boldsymbol{mean_{cluster}}) + \frac{\mathbf{1}}{\mathbf{2}} \cdot (\boldsymbol{SD_{BKG}} + \boldsymbol{SD_{cluster}}) \tag{1}$$

After thresholding, morphological opening and closing were conducted to extract the area of the clustered glia above the threshold. The extracted area was then used to estimate the number of cells in the region, simply by dividing the area by 2000 μm^2 and rounding the result.

The root-mean-square error (RMSE) was used to evaluate the area-based counting. For each color of glia, we calculate the RMSE as below.

$$\boldsymbol{RMSE} = \sqrt{\frac{\sum_{i=1}^{M}(\hat{n}_i - n_i)^2}{M}} \tag{2}$$

$\hat{n}$ is the estimate number of glia in the cluster. n is the ground truth number. M is the total number of detected glia clusters.

3.7 Evaluation metrics

In object detection tasks [29], precision is defined as: the number of correct predictions, also named as true positives (TPs), divided by the number of all predictions. Higher precision means fewer false positives (FPs).

$$\boldsymbol{Precision} = \frac{\boldsymbol{TP}}{\boldsymbol{TP} + \boldsymbol{FP}} \tag{3}$$

Recall is defined as the number of correct predictions divided by the number of all ground-truth positives i.e., including false negatives (FNs). Recall illustrates the sensitivity of a model, and high recall value indicates low number of false negatives.

$$\boldsymbol{Recall} = \frac{\boldsymbol{TP}}{\boldsymbol{TP} + \boldsymbol{FN}} \tag{4}$$

To comprehensively measure these two aspects of a model, F-score (F_1) is utilized [30]. Similar to precision and recall, the F-score values are ranging from 0 to 1. Value of 1 indicates that the predictions and ground-truth annotations are identical.

$$\boldsymbol{F_1} = \frac{\boldsymbol{2 \cdot Precision \cdot Recall}}{\boldsymbol{Precision + Recall}} \tag{5}$$

To evaluate the performance of an object detection model, we followed the average precision measure [31], in terms of the relative overlap of the bounding boxes. During the inference stage, an object detection model will output per prediction the coordinates of a bounding box, a classification label, and a confidence score. For each class, the output predictions are first ranked according to their corresponding confidence scores. Then a set of precision and recall will be calculated following the confidence rank. A prediction is true if the intersection over union (IoU) of the bounding box of the prediction and an undetected object is above 0.5, where the IoU is defined as the area of intersection divided by the area of union given two bounding boxes.

Starting from the prediction with the highest confidence score, the first pair of precision and recall are calculated based on this prediction only. Then adding the prediction with the second highest confidence score, the second pair of precision and recall are calculated considering top two predictions, and so on at each rank. Finally, a set of precision and recall values are obtained, and an initial precision-recall curve is depicted in which each point is a pair of precision and recall. For each recall value r, the precision value p_r is replaced by the maximum precision value $p_{r'}$ whose corresponding recall value is no less than the original recall value ($r' \geq r$). The AP of the class is obtained by calculating the area under the adjusted precision-recall curve. For multiclass detection, the overall AP is an average of AP values across all classes.

3.8 Statistical analysis

Unpaired t-tests assuming unequal variances were performed in Microsoft Excel.

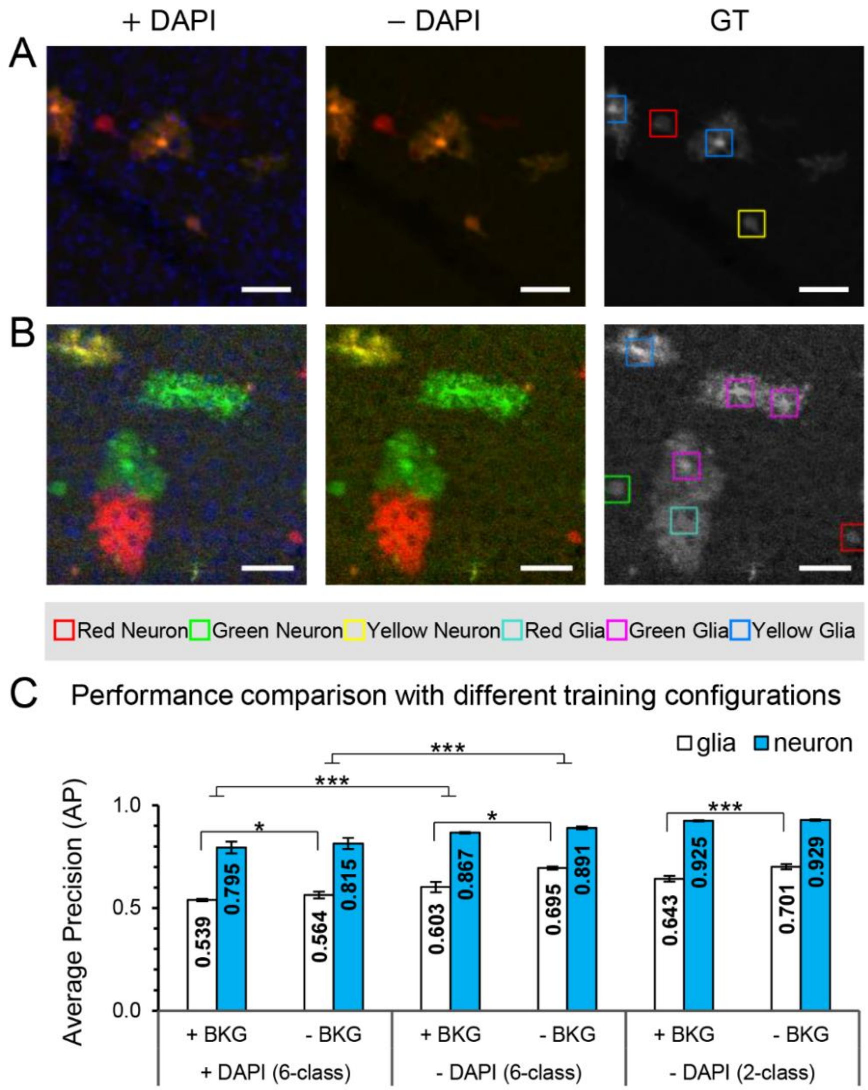
4 Results

4.1 Training configuration experiments

A RetinaNet model was trained to detect six classes of individual cells: green glia, yellow glia, red glia, green neuron, yellow neuron, and red neuron. In addition to the conventional parameters (e.g., optimization hyper parameters) our training approach was adapted to account for the unique properties of our datasets. We designed three different training configurations and tested their effects on the network's performance: (*i*) Adding a DAPI channel, which reveals stained cell nuclei, as an input to the network. (*ii*) Training the network without background tiles. (*iii*) Adopting a color-independent detection approach to classify cells. In all the comparisons between the models, an average precision (AP) measure was used [31].

Both the slide scanner (Figure 2A) and confocal (Figure 2B) datasets were acquired on three color channels: GFP (green), RFP (red) and DAPI (blue). While GFP and RFP reporters are genetically expressed in a subset of cells permanently, the DAPI stain binds to DNA inside all the cells in a sample. Therefore, an experiment of evaluating the network's performance with or without a DAPI channel was conducted. Training without the DAPI channel significantly improved the model performance of six-class detection from 0.667 ± 0.017 to 0.735 ± 0.013 (mean ± SD, $p < 0.005$, unpaired t-test, $n = 3$; Figure 2C). It may be possible that the lack of cell-specificity inherent to the DAPI channel obscured successful detection of the red and green channels.

Figure 2. Training configurations that influence the performance of a single RetinaNet model in individual cell detection. (**A, B)** Representative images from the slide scanner and the confocal fluorescence microscope (CFM) respectively, with the corresponding ground truth (GT) annotations. The GT annotations are on grayscale images for clearer display of the bounding boxes. Scale bars, 50 µm. (**C**) Average precision (mean ± SD, n = 3; *, $p < 0.05$; ***, $p < 0.005$; unpaired t-test) of trained models with different training configurations. Configurations include training: (*i*) with and without the DAPI channel. (*ii*) With and without pure-background (BKG) image patches (i.e., no target cells in the image patches). (*iii*) On six classes (red/green/yellow neuron, and red/green/yellow glia) versus two color-independent classes (neuron, glia). Training with the DAPI channel and BKG fail to improve the performance and degrade the average precision. Two-class detection shows better performance on neurons without color classification. Please note that when the significance line ends with inverted T, it shows significance between the average of the two classes.



+ DAPI
– DAPI
GT
A
B
Red Neuron
Green Neuron
Yellow Neuron
Red Glia
Green Glia
Yellow Glia
C
Performance comparison with different training configurations
glia
neuron
Average Precision (AP)
1.0
0.5
0.0
0.539
0.795
0.564
0.815
0.603
0.867
0.695
0.891
0.643
0.925
0.701
0.929
*
***
+ BKG
- BKG
+ DAPI (6-class)
- DAPI (6-class)
- DAPI (2-class)

When generating the training data, a full brain section image was cropped into small image patches. As a result, in a sparse sample, there were many pure-background image patches (i.e., patches without any cells), whereas the target cells were distributed sparsely across the tissue. For instance, in a sparse section there were 70 image patches of pure background compared to 33 image patches that contained target cells (Figure 1C). Such pure-background image patches were considered as negatives during training, but failed to improve and even slightly degraded the performance of the model (Figure 2C). Therefore, we decided to train without pure background patches hereafter.

Next, we trained a RetinaNet model to detect only two classes, neurons and glia regardless of their MADM colors. Given the same training data, reducing the number of classes led to better performance, mainly for the neuron class from AP of 0.891 ± 0.007 to 0.929 ± 0.003 (mean ± SD, $p < 0.05$, unpaired t-test, $n = 3$; Figure 2C). This improvement in performance was likely to result from the imbalance of different colors in the training data. It should be noted that for comparison with the two class models, we averaged the AP values of three colors for either neurons or glia in the six-class model.

4.2 Data augmentation for improved detection of individual cells

To match the performance of the six-class model to the two-class model, we aimed to eliminate color dependent biases in the training data. Toward this end, we doubled the size of the training set by adding a color swapped version of the original (Figure 3). We used the color swap approach since all the neurons independent on their colors looked very similar, and this was also true for glial cells. We also found that the saturation of the input data might be color-dependent, hence we multiplied the original image by a constant and used a ceiling function to emulate saturation (see methods section). After training with each type of augmentation condition, we found that data augmentation improved the detection of individual cells. Utilizing all types of data augmentation together led to the best performance of individual cell detection with AP of 0.860 ± 0.006 (mean ± SD, $p < 0.005$, unpaired t-test, $n = 3$; Figure 3C), which exceeded the performance of the two-class model with AP of 0.815 ± 0.005 (mean ± SD, $n = 3$) while maintaining the ability to distinguish MADM colors. This result reiterated the importance of tailoring the data augmentation to the unique properties of the dataset. Moreover, numerical experiments were performed to evaluate the model’s performance with different amount of training data (Figure 3D). Figure 3D shows that as the amount of training data increased, a plateau of performance was reached.

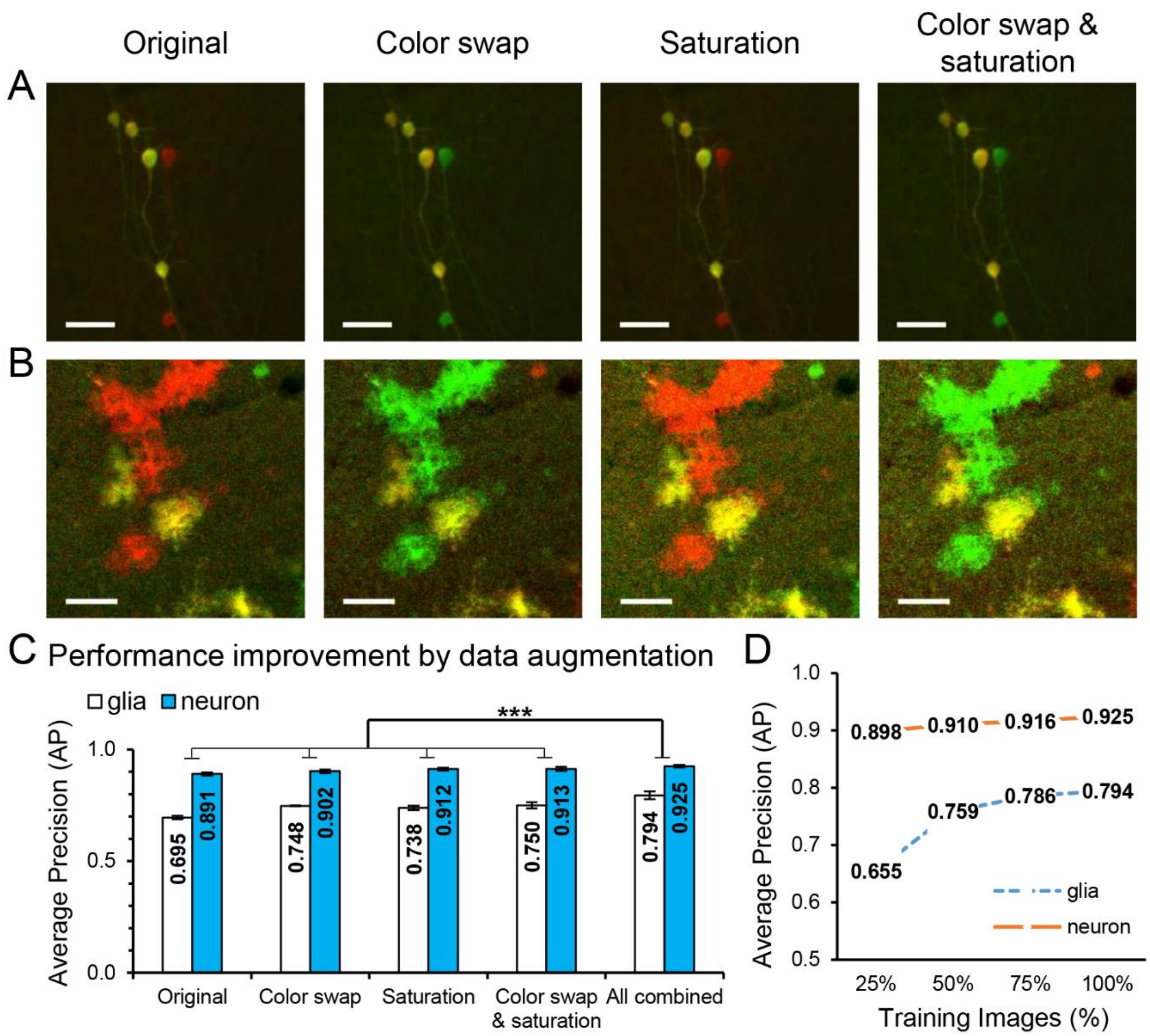


Figure 3. Data augmentation by color swap and saturation improve individual cell detection in a single RetinaNet model. (**A, B)** Slide scanner and confocal images respectively with data augmentation including color swap and/or saturation. Scale bars, 50 µm. (**C**) Average precision results (mean ± SD, n = 3; ***, $p < 0.005$; unpaired t-test) from the different augmentation conditions. Utilizing each type of data augmentation provided similar results. Harnessing all three augmentations together led to the best performance in both neuron and glia detection. (D) Average precision results with respect to proportion of training images. A plateau of performance was reached when increasing the amount of training data.

4.3 A single RetinaNet model reached an average AP of 0.90 on individual cells

After fine-tuning a RetinaNet model using the data augmentation techniques presented in Figure 3, we tested the model performance on an independent test set i.e., the test set was derived from different mouse brain samples from those used for training. Table 1 summarizes the AP of individual cell detection across six classes and after averaging the test results from three independent networks, which were trained on the same dataset. Detection of neurons reached an AP of 0.943 ± 0.005 (mean ± SD, n = 3), while the glial cells reached an AP of 0.857 ± 0.002 (mean ± SD, n = 3). The precision-recall curves of the best model are shown in Figure S1. Moreover, the method was evaluated using 5-fold cross validation where an AP of 0.90 was obtained (Table 2). For completeness, we also compared the results of the two-class model (neurons and glia regardless of their MADM colors) after data augmentation with the six-class model. The two-class detection model had an average precision of 0.952 ± 0.005 (mean ± SD, n = 3) for neurons and 0.863 ± 0.011 (mean ± SD, n = 3) for glial cells. These are comparable results to the six-class detection model, but this comparison is doing a disservice to the six-class model, since on top of detecting neurons and glia, it also had to classify the cells by color.

To further improve these results, we studied cases in which the model made correct (Figure 4A) and incorrect (Figure 4B) predictions. Based on observations, we reasoned that the larger inherent variability in the glial cells' morphology, and their tendency to form dense clusters underlay the better performance achieved by RetinaNet in identifying neurons versus glia. This discrepancy is exacerbated when the glia clusters are saturated as a result of higher density of glial processes per pixel. Therefore, we hypothesized that independent detection of glia clusters may improve the performance of the model.

Table 1. Average precision results (mean ± SD, n = 3) across six classes in individual cell detection using a single RetinaNet model.

	Neuron	Glia	Average
Yellow	0.950 ± 0.005	0.876 ± 0.011	0.913 ± 0.008
Green	0.955 ± 0.006	0.849 ± 0.017	0.902 ± 0.007
Red	0.925 ± 0.004	0.847 ± 0.001	0.886 ± 0.002
Average	**0.943 ± 0.005**	**0.857 ± 0.002**	**0.900 ± 0.001**

Table 2. Average precision results (mean ± SD, n = 5) of 5-fold cross validation across six classes in individual cell detection using a single RetinaNet model.

	Neuron	Glia	Average
Yellow	0.932 ± 0.028	0.857 ± 0.023	0.895 ± 0.022
Green	0.953 ± 0.030	0.850 ± 0.112	0.902 ± 0.052
Red	0.947 ± 0.013	0.880 ± 0.012	0.914 ± 0.008
Average	**0.944 ± 0.018**	**0.863 ± 0.037**	**0.903 ± 0.016**

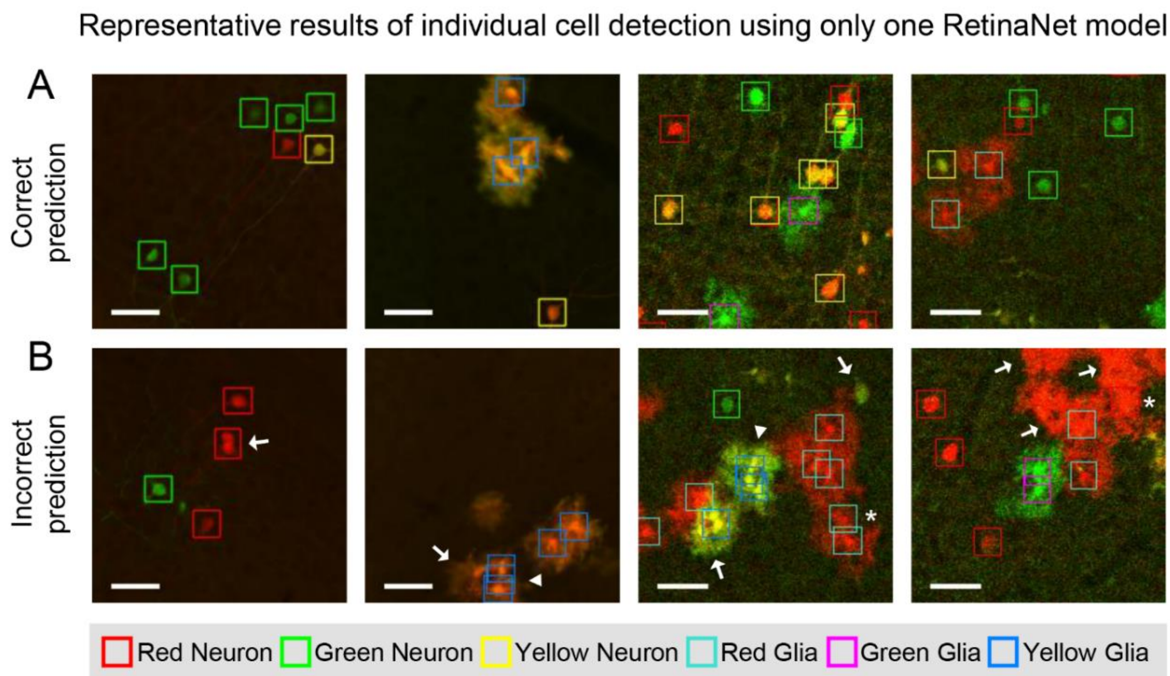


Figure 4. Representative model-detected neurons and glia. (**A)** Examples of correct predictions on images from both the slide scanner (first and second images) and confocal (third and fourth images). (**B)** Examples of incorrect predictions from the trained model. Three main error types including miss detection (arrow), redundant detection (arrowhead) and false detection (asterisk) are marked in the images. Glia clusters and their saturation were the main factors that caused false predictions and miss detections. Threshold of confidence was 0.5 in all images. Scale bars, 50 µm.

4.4 Merging results of two RetinaNet models improved detection and classification of glia

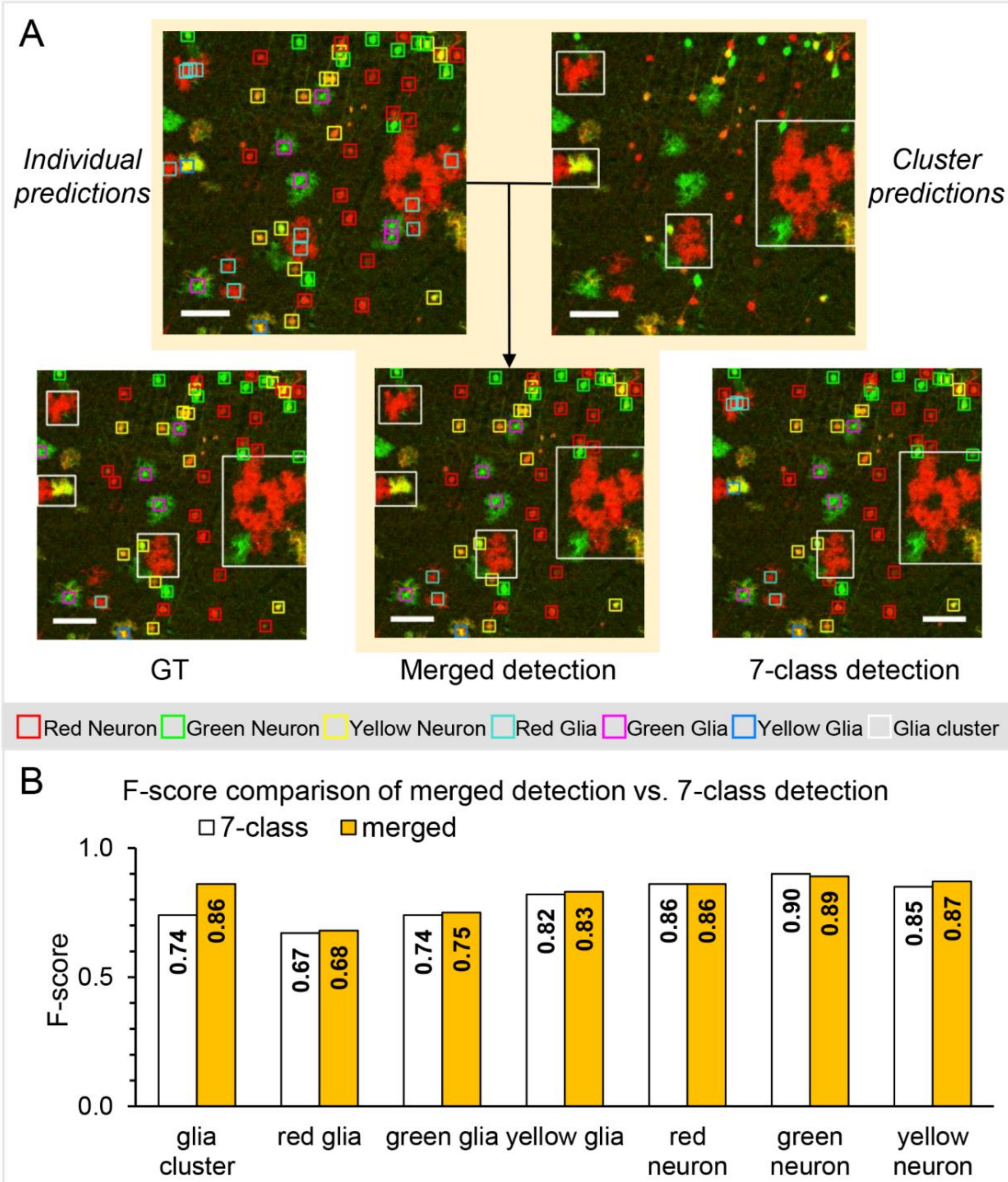
We found that the detection of individual glial cells in a cluster was difficult for the network, as it is for a manual annotator. To address this issue, we first made an attempt of training a single RetinaNet model to detect seven classes: the previous six classes and an additional glia cluster class. In the training data of seven-class detection model, glia clusters with various sizes were annotated separately from isolated individual glia. Overall, we found that this approach failed to perform well in detection of glia clusters.

Next, two RetinaNet models were trained (Figure 1E), one for individual cells (see Table 1 for AP results), and one to solely detect glia clusters. The AP of glia cluster detection was 0.76 on an independent test set. Lower AP is expected for glia clusters due to variability in the glia clusters in terms of size and morphology, which hinders consistent annotation of bounding boxes even for a manual annotator. Then the results from the two models were integrated using a rule-based merging process. The three rules are: (*i*) keep all detected clusters with a confidence score above 0.5; (*ii*) keep glia clusters with confidence above 0.3 that have overlapping individual glial cells and remove these individual cells. In such cases, the individual cells provide evidence that increases our confidence that a cluster is present. (*iii*) Eliminate redundant clusters – i.e., when more than half of the bounding boxes for nearby clusters overlap.

Examples of the merging process are shown (Figure 5A, and Figure S2) with the corresponding ground truth and the seven-class detection results. To evaluate the results, we merged ground truth annotations according to the same aforementioned rules and calculated the F-score for each class on the independent test set. The merged results consistently showed higher F-scores in glial classes than using the seven-class detection network, especially on the saturated images (Figure 5B). Particularly, the F-score for detecting glia cluster was higher using the two RetinaNet models versus the seven-class network (0.86 and 0.74, respectively).

To determine the number of glia within a detected glia cluster, we used area-based cell counting to estimate the number of individual cells. Cell regions were extracted by thresholding and morphological operations, and the number of individual cells was then estimated according to the area of the extracted cell masks (see methods section and Figure S3). The root-mean-square error (RMSE) was used to evaluate the counting results of the detected clusters. Counting results obtained for 23 test glia clusters, show that the results of merging the two RetinaNet models were better than six-class detection with a RMSE of 0.59 compared to 1.43 in red glia. Please note, that the red glia was the most common type of cluster in this specific dataset. For green and yellow glia, the counting in glia clusters by merging the two RetinaNet models had RMSEs of 0.36 and 0.69 respectively, which was comparable to the six-class detection with RMSEs of 0.21 and 0.36 respectively.

Figure 5. Combining predictions from two RetinaNet models enhances performance. (**A)** An example of merging results compared to the ground truth (GT) annotations. As shown in Figure 1E, two RetinaNet models were trained separately: one for individual cell detection and one for glia cluster detection. Predictions from both the trained individual cell-detection models and the trained glia cluster-detection model were then merged to assess performance. For comparison, an additional RetinaNet model was trained to detect seven classes simultaneously (glia cluster, red/green/yellow glia, and red/green/yellow neuron). Predictions on the same image patch with confidence scores above 0.5 are shown. Scale bars, 100 µm. **(B)** F-score distributions for merged detection and seven-class detection on the test dataset. F-score is a comprehensive measure of accuracy combining precision and recall. F-score in the glia cluster improved significantly by merging predictions from the two RetinaNet models.


A
Individual predictions
Cluster predictions
GT
Merged detection
7-class detection
Red Neuron
Green Neuron
Yellow Neuron
Red Glia
Green Glia
Yellow Glia
Glia cluster
B
F-score comparison of merged detection vs. 7-class detection
7-class
merged
F-score
1.0
0.5
0.0
0.74
0.86
0.67
0.68
0.74
0.75
0.82
0.83
0.86
0.86
0.90
0.89
0.85
0.87
glia cluster
red glia
green glia
yellow glia
red neuron
green neuron
yellow neuron

5 Discussion

In this study, we developed an automatic cell detection workflow that was applied to images obtained from MADM-labeled mouse brain sections. Our workflow achieved an overall AP of 0.90 ± 0.001 (mean ± SD, n = 3) for individual cell detection, with an AP of 0.943 ± 0.005 and 0.857 ± 0.002 for individual neurons and glial cells, respectively. We also trained YOLOv3 and SSD models instead of RetinaNet to detect individual cells, and the best AP results that we achieved were 0.787 and 0.793, respectively. Please note that we did not spend extensive amount of time optimizing the models as compared to RetinaNet. To detect dense and slightly saturated glia clusters, we incorporated an additional RetinaNet model. This approach showed superior performance in comparison with a more traditional approach, i.e., a single RetinaNet model with seven classes (individual cells plus glia clusters). We also presented a novel data augmentation method that was used to compensate for color-, intensity-, and saturation-dependent biases in the dataset due to the investigated genotype and acquisition conditions. To the best of our knowledge, this is the first paper to integrate multiple fluorescence channels (except the DAPI channel) and to use the RetinaNet model for color classification. We believe that the presented approach could be used in multiple tissue preparations and in quantification of various structures with double stain. This is especially appealing as the training of detection networks is fast and relatively simple.

Although in the current stage of its development our workflow has shown great promise, it still faces several limitations. First, deep learning approaches are black box methods, i.e., it is difficult to explain the relationship between the input and the output, and therefore improve their performance in scenarios where they fail. It will be interesting to explore and utilize explainable artificial intelligence (XAI) methods, which can be interpreted by humans [32–34]. However, XAI still faces many challenges to obtain explainability in deep learning models [32]. Second, the presented approach did not explore the multiscale capability of the RetinaNet model, i.e., detection under different magnification conditions. Please note that magnification in this context is not related to structural hierarchies or different orders of structure [35,36], magnification relates to the ability of a microscope to produce a larger image of the object, relative to its actual size. This multiscale capability provides flexibility to integrate datasets that were acquired with different magnifications. Here, the multiscale capability of the RetinaNet model was not tested, since the acquired datasets had similar magnification. Future work will introduce more variability to our datasets, by including data with different magnifications, and from additional imaging modalities such as light-sheet fluorescence microscopy. Third, detecting individual glial cells using a 10 × magnification (~ 0.3 numerical aperture) in a cluster is challenging, not only for a machine but also to a manual annotator. Therefore, other than counting cells within the cluster using area-based cell counting, we will reimage the dense and challenging regions with higher resolution e.g., 40 × magnification (~ 0.8 numerical aperture), and high-dynamic-range acquisition. Additionally, we plan to integrate an object detection network into a microscope in the near future. This will enable the real time detection of problematic regions, and in turn allow for local reimaging at higher magnifications whenever further resolution of these ambiguous cases in datasets is needed. Such need-based approaches of automatic acquisition will translate into an efficient way to utilize expensive microscopes and to compress the raw dataset sizes.

Last, the double marker approach in our case provided the cell genotype and the markers/colors should spatially overlap. However, in other applications such as diagnosis of non-small cell lung cancer [37] the intracellular localization of the double markers might reveal important information. In these cases, segmentation of the markers will be required, thus revealing a potential limitation for object detection approaches.

REFERENCES

1. Ruprecht V, Monzo P, Ravasio A, Yue Z, Makhija E, Strale PO, et al. How cells respond to environmental cues – insights from bio-functionalized substrates. J Cell Sci. 2017 Jan 1;130(1):51–61.
2. Suárez LE, Markello RD, Betzel RF, Misic B. Linking Structure and Function in Macroscale Brain Networks. Trends in Cognitive Sciences. 2020 Apr 1;24(4):302–15.
3. Chung K, Wallace J, Kim SY, Kalyanasundaram S, Andalman AS, Davidson TJ, et al. Structural and molecular interrogation of intact biological systems. Nature. 2013 May 16;497(7449):332–7.
4. Chung K, Deisseroth K. CLARITY for mapping the nervous system. Nat Methods. 2013 Jun;10(6):508–13.
5. Moatti A, Moatti A, Cai Y, Cai Y, Li C, Li C, et al. Three-dimensional imaging of intact porcine cochlea using tissue clearing and custom-built light-sheet microscopy. Biomed Opt Express, BOE. 2020 Nov 1;11(11):6181–96.
6. Zong H, Espinosa JS, Su HH, Muzumdar MD, Luo L. Mosaic Analysis with Double Markers in Mice. Cell. 2005 May;121(3):479–92.
7. Gao P, Postiglione MP, Krieger TG, Hernandez L, Wang C, Han Z, et al. Deterministic progenitor behavior and unitary production of neurons in the neocortex. Cell. 2014 Nov 6;159(4):775–88.
8. Johnson CA, Ghashghaei HT. Sp2 regulates late neurogenic but not early expansive divisions of neural stem cells underlying population growth in the mouse cortex. Development. 2020 Feb 21;147(4).

9. Laukoter S, Beattie R, Pauler FM, Amberg N, Nakayama KI, Hippenmeyer S. Imprinted Cdkn1c genomic locus cell-autonomously promotes cell survival in cerebral cortex development. Nat Commun. 2020 Jan 10;11(1):195.

10. Beattie R, Postiglione MP, Burnett LE, Laukoter S, Streicher C, Pauler FM, et al. Mosaic Analysis with Double Markers Reveals Distinct Sequential Functions of Lgl1 in Neural Stem Cells. Neuron. 2017 May 3;94(3):517-533.e3.

11. Zhang X, Mennicke CV, Xiao G, Beattie R, Haider MA, Hippenmeyer S, et al. Clonal Analysis of Gliogenesis in the Cerebral Cortex Reveals Stochastic Expansion of Glia and Cell Autonomous Responses to Egfr Dosage. Cells. 2020 Dec 11;9(12).

12. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015 May;521(7553):436–44.

13. Voulodimos A, Doulamis N, Doulamis A, Protopapadakis E. Deep Learning for Computer Vision: A Brief Review. Comput Intell Neurosci. 2018;2018:7068349.

14. Bayramoglu N, Kannala J, Heikkila J. Deep learning for magnification independent breast cancer histopathology image classification. In: 2016 23rd International Conference on Pattern Recognition (ICPR) [Internet]. Cancun: IEEE; 2016 [cited 2021 Feb 16]. p. 2440–5. Available from: http://ieeexplore.ieee.org/document/7900002/

15. Todorov MI, Paetzold JC, Schoppe O, Tetteh G, Shit S, Efremov V, et al. Machine learning analysis of whole mouse brain vasculature. Nat Methods. 2020 Apr;17(4):442–9.

16. Jiang Z, Liu X, Yan Z, Gu W, Jiang J. Improved detection performance in blood cell count by an attention-guided deep learning method. OSA Continuum. 2021 Feb 15;4(2):323.

17. Huang J, Rathod V, Sun C, Zhu M, Korattikara A, Fathi A, et al. Speed/Accuracy Trade-Offs for Modern Convolutional Object Detectors. In: 2017 IEEE Conference on Computer

Vision and Pattern Recognition (CVPR) [Internet]. Honolulu, HI: IEEE; 2017 [cited 2019 Dec 9]. p. 3296–7. Available from: http://ieeexplore.ieee.org/document/8099834/

18. Zou Z, Shi Z, Guo Y, Ye J. Object Detection in 20 Years: A Survey. arXiv:190505055 [cs] [Internet]. 2019 May 15 [cited 2019 Dec 9]; Available from: http://arxiv.org/abs/1905.05055

19. Lin TY, Goyal P, Girshick R, He K, Dollár P. Focal Loss for Dense Object Detection. arXiv:170802002 [cs] [Internet]. 2018 Feb 7 [cited 2019 Dec 9]; Available from: http://arxiv.org/abs/1708.02002

20. Redmon J, Divvala S, Girshick R, Farhadi A. You Only Look Once: Unified, Real-Time Object Detection. arXiv:150602640 [cs] [Internet]. 2016 May 9 [cited 2019 Dec 9]; Available from: http://arxiv.org/abs/1506.02640

21. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, et al. SSD: Single Shot MultiBox Detector. arXiv:151202325 [cs]. 2016;9905:21–37.

22. Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Trans Pattern Anal Mach Intell. 2017 Jun 1;39(6):1137–49.

23. Redmon J, Farhadi A. YOLOv3: An Incremental Improvement. arXiv:180402767 [cs] [Internet]. 2018 Apr 8 [cited 2021 Jun 25]; Available from: http://arxiv.org/abs/1804.02767

24. Waithe D, Brown JM, Reglinski K, Diez-Sevilla I, Roberts D, Eggeling C. Object detection networks and augmented reality for cellular detection in fluorescence microscopy. Journal of Cell Biology. 2020 Oct 5;219(10):e201903166.

25. Hippenmeyer S, Youn YH, Moon HM, Miyamichi K, Zong H, Wynshaw-Boris A, et al. Genetic mosaic dissection of Lis1 and Ndel1 in neuronal migration. Neuron. 2010 Nov 18;68(4):695–709.

26. Liang H, Xiao G, Yin H, Hippenmeyer S, Horowitz JM, Ghashghaei HT. Neural development is dependent on the function of specificity protein 2 in cell cycle progression. Development. 2013 Feb 1;140(3):552–61.

27. Berg S, Kutra D, Kroeger T, Straehle CN, Kausler BX, Haubold C, et al. ilastik: interactive machine learning for (bio)image analysis. Nat Methods. 2019 Dec;16(12):1226–32.

28. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. Nat Methods. 2012 Jul;9(7):676–82.

29. Padilla R, Netto SL, Silva EAB da. A Survey on Performance Metrics for Object-Detection Algorithms. In: 2020 International Conference on Systems, Signals and Image Processing (IWSSIP). 2020. p. 237–42.

30. Hossin M, Sulaiman MN. A Review on Evaluation Metrics for Data Classification Evaluations. International Journal of Data Mining & Knowledge Management Process. 2015 Mar 31;5:01–11.

31. Everingham M, Eslami SMA, Van Gool L, Williams CKI, Winn J, Zisserman A. The Pascal Visual Object Classes Challenge: A Retrospective. Int J Comput Vis. 2015 Jan;111(1):98–136.

32. Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion. 2020 Jun 1;58:82–115.

33. Anguita-Ruiz A, Segura-Delgado A, Alcalá R, Aguilera CM, Alcalá-Fdez J. eXplainable Artificial Intelligence (XAI) for the identification of biologically relevant gene expression patterns in longitudinal human studies, insights from obesity research. PLoS Comput Biol. 2020 Apr;16(4):e1007792.

34. Alonso JM, Casalino G. Explainable Artificial Intelligence for Human-Centric Data Analysis in Virtual Learning Environments. In: Burgos D, Cimitile M, Ducange P, Pecori R, Picerno P, Raviolo P, et al., editors. Higher Education Learning Methodologies and Technologies Online [Internet]. Cham: Springer International Publishing; 2019 [cited 2021 Jun 30]. p. 125–38. (Communications in Computer and Information Science; vol. 1091). Available from: http://link.springer.com/10.1007/978-3-030-31284-8_10

35. Kayser K, Borkenfeld S, Djenouni A, Kayser G. Texture and object related image analysis in microscopic images. Diagnostic Pathology [Internet]. 2015; Available from: http://www.diagnosticpathology.eu/content/index.php/dpath/article/view/14

36. Kayser K, Borkenfeld S, Kayser G. Digital Image Content and Context Information in Tissue-based Diagnosis. Diagnostic Pathology [Internet]. 2018;4(1). Available from: http://www.diagnosticpathology.eu/content/index.php/dpath/article/view/269

37. Antfang C, Csanadi A, Kayser G. Expression of PTEN and pAKT in Non-Small Cell Lung Cancer. Diagnostic Pathology [Internet]. 2019;5(1). Available from: http://www.diagnosticpathology.eu/content/index.php/dpath/article/view/272

CHAPTER 3: COMBINe enables automated detection and classification of neurons and astrocytes in tissue cleared mouse brains

Yuheng Cai[1,2], Xuying Zhang[3], Chen Li[1,2], H. Troy Ghashghaei[3], Alon Greenbaum[1,2,4,5*]

[1] Joint Department of Biomedical Engineering, North Carolina State University and University of North Carolina at Chapel Hill, Raleigh, North Carolina, United States of America

[2] Comparative Medicine Institute, North Carolina State University, Raleigh, North Carolina, United States of America

[3] Department of Molecular Biomedical Sciences, North Carolina State University, Raleigh, North Carolina, United States of America

[4] Bioinformatics Research Center, North Carolina State University, Raleigh, North Carolina, United States of America

[5] Lead contact

[*] Correspondence: greenbaum@ncsu.edu

Cai, Y., Zhang, X., Li, C., Ghashghaei, H. T., & Greenbaum, A. (2023). COMBINe enables automated detection and classification of neurons and astrocytes in tissue-cleared mouse brains. *Cell Reports Methods*, *3*(4). https://doi.org/10.1016/j.crmeth.2023.100454

1 Highlights

- COMBINe performs automated cell detection based on morphology and color

- COMBINe accurately detects out-of-focus cells in tissue cleared brains
- COMBINe overcomes slight misalignments between color channels
- Hierarchical analysis reveals regional effect of sparse *Egfr* deletion on astrocytes

2 eTOC Blurb

Cai et al. present a workflow called COMBINe to automatically detect and quantify all labeled cells in tissue cleared mouse brains and apply it to study the effects of sparse Epidermal Growth Factor Receptor deletion. COMBINe is adaptable for various neuroscience projects that require whole-brain quantification and analysis.

3 Summary

Tissue clearing renders entire organs transparent to accelerate whole tissue imaging for example with light sheet fluorescence microscopy. Yet, challenges remain in analyzing the large resulting 3D datasets that consist of terabytes of images and information on millions of labeled cells. Previous work has established pipelines for automated analysis of tissue cleared mouse brains, but the focus there was on single color channels and/or detection of nuclear localized signals in relatively low-resolution images. Here, we present an automated workflow (COMBINe, Cell detectiOn in Mouse BraIN) to map sparsely labeled neurons and astrocytes in genetically distinct mouse forebrains using Mosaic Analysis with Double Markers (MADM). COMBINe blends modules from multiple pipelines with RetinaNet at its core. We quantitatively analyzed the regional and subregional effects of MADM-based deletion of the Epidermal Growth Factor Receptor (EGFR) on neuronal and astrocyte populations in the mouse forebrain.

4 Keywords

Deep Learning, MADM, Mouse Brain, Tissue Clearing, Cell Detection, Neuron, Astrocyte

5 Motivation

Tissue clearing has been extensively used to explore cellular organizations in intact organs. By imaging entire mouse brains with cellular resolution, the resulting datasets contain terabytes of images to process and millions of labeled cells that need to be counted and classified based on morphology and color. Therefore, automation of data processing and analysis is in urgent need. We present a workflow (COMBINe) to automatically locate and classify labeled cells in such 3D datasets and to quantitatively analyze the regional effects after registration to the referenced Allen Brain Atlas. We applied the approach to study the effects of sparse Epidermal Growth Factor Receptor deletion on gliogenesis.

6 Introduction

Recent advances in tissue clearing methods have enabled three-dimensional (3D) imaging of intact mammalian organs and entire small organisms while preserving comprehensive structural and cellular information[1–7]. With light sheet fluorescence microscopy (LSFM), high-speed imaging of cleared and immunolabeled tissues can be achieved[8–11]. Yet, challenges remain in analyzing the large resulting 3D datasets with terabytes of images and information on millions of labeled cells. Hence, the automation of data processing and cell detection is paramount for high throughput cellular profiling of cleared tissues.

Existing pipelines for automated analysis of tissue cleared mouse brains include BCFind[12,13], ClearMap[5], CUBIC[14], and NuMorph[15]. BCFind achieves fully automated detection of neurons in 3D mouse brains by integrating a U-Net model for semantic deconvolution and a blob detection algorithm for cell identification. ClearMap presents an immunolabeling-enabled tissue clearing method with superior efficiency in data processing, accomplishing automated cell counting, dataset registration, and statistical analysis. The datasets are analyzed using traditional image processing routines (e.g., morphological operations). CUBIC incorporates improved clearing, imaging, and cell-nucleus-detection protocols, with the ability to quantify cells of entire adult mouse brains. NuMorph is an advanced end-to-end data processing tool for accurate cell-type quantification within the mouse cortex, which reaches high precision in nuclei detection. The previous pipelines achieved high efficacy, yet they focused on single color channels and/or detection of nuclear signals, in relatively low-resolution images. As such, the existing pipelines fail to capture cases where 1) whole cells or cell membranes rather than nuclei are stained, 2) colocalization of color channels is crucial, and 3) the cells can be classified based on their morphology which requires high-resolution imaging. These requirements are broadly paramount in neuroscience research, especially those that require in situ analysis of multiple genetic markers as in Confetti mice[16] or Mosaic Analysis with Double Markers (MADM)[17–21]. Here we addressed this critical gap by building on the strengths of existing pipelines while developing an automated workflow that was tested in MADM datasets. In these datasets the fluorescent expression in neurons and glia report on their distinct genotypes.

In the core of our workflow is deep learning, a data driven approach. Recent adaptation of deep learning[22] in biomedical studies has enabled automated and accelerated image processing[23–25]. In combination with tissue clearing and 3D microscopy, deep learning has been employed in investigation of cancer metastasis in mice[26], segmentation of mouse brain vasculature[27], and detection of crown-like structures in adipose tissues[28]. Inspired with these studies, we surveyed the deep learning literature, and employed a model (RetinaNet) that shows superior performance in dense object detection and applied it to a neurodevelopmental problem[29–31].

With RetinaNet in its core, we developed an automated workflow we refer to as COMBINe to map labeled neurons and astrocytes in the forebrains of three genetically distinct MADM mice (Figure 1). We used a conditional mouse allele for the *Epidermal growth factor receptor* (*Egfr*) in combination with MADM alleles. Using the established *Emx1^{cre}* line to restrict recombination to the dorsal telencephalon (cerebral cortices and the hippocampal formation), three genotypes were obtained: *Emx:MADM:+/+, Emx:MADM:F/+, and Emx:MADM:F/F*[32]. We have recently reported on variable effects on specification and differentiation of MADM glia in cerebral cortices of these mice, which has led to the developmental model that EGFR-dependent and EGFR-independent perinatal gliogenesis regulate forebrain development cooperatively[32]. Here we used the *Egfr* MADM model to test COMBINe. Using tissue clearing and a custom-built LSFM that changes its properties based on the sample, we generated high-quality MADM datasets with cellular resolutions sufficient to morphologically differentiate astrocytes from neurons. Employing RetinaNet, 0.5 million MADM cells across six classes (yellow, red, green: neurons and astrocytes) were located and counted from each MADM forebrain hemisphere. Taking advantage of available packages, stitched volumes and detected cells were aligned to annotations provided by the Allen Brain Atlas (ABA) for regional registration and analysis[33]. Significant variations in cell populations corresponding to the

MADM genotypes were confirmed matching our findings from laborious manual counting and registrations.

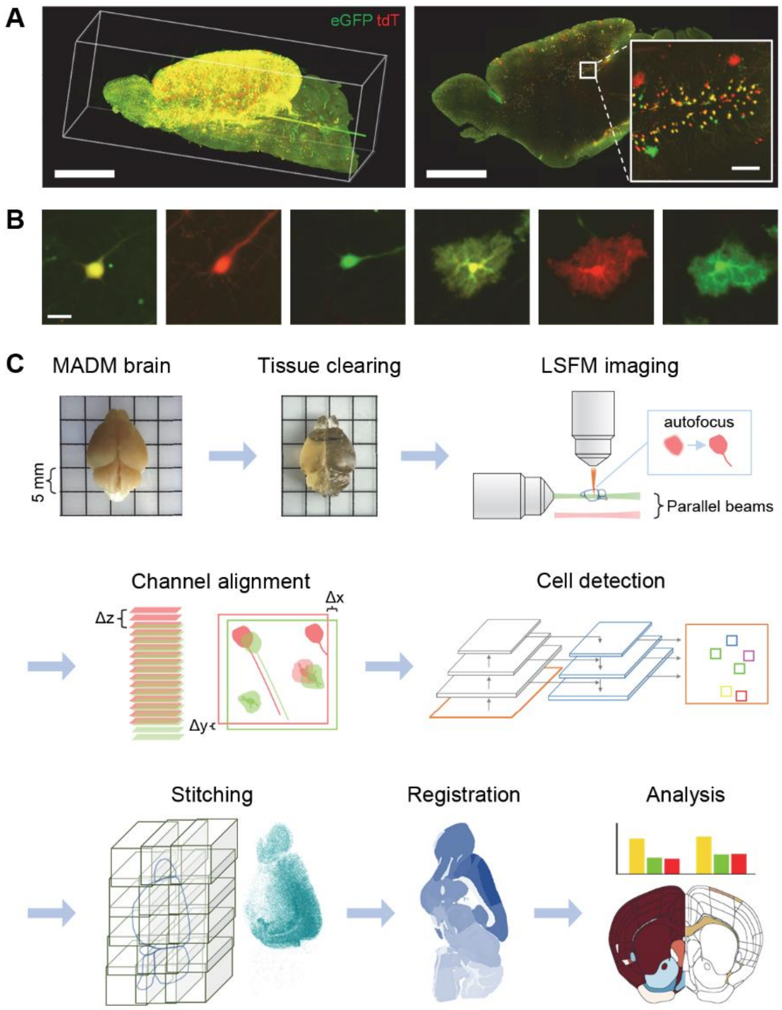


Figure 1. Automated cell detection in the cleared MADM mouse forebrain. (A) Left: Maximum intensity projection of fluorescence signals from a hemisphere of an *Emx:MADM:+/+* forebrain labeled with antibodies to enhance the endogenous eGFP (green) and tdTomato (tdT, red) signals (scale bar, 3 mm). Right: A virtual 30 µm thick slice from the same cleared sample (scale bar, 2.5 mm) with a zoomed-in area (scale bar, 100 µm). eGFP – enhanced green fluorescent protein, tdT – a red fluorescent protein. **(B)** Magnified images of three isolated MADM-labeled neurons and astrocytes expressing eGFP (green), tdT (red), or both (yellow) MADM reporters. Scale bar, 25 µm. **(C)** The proposed COMBINe workflow for MADM cell mapping (see also Figure S1).

7 Results

7.1 Automated cell detection in cleared MADM mouse forebrain

We combined the *Emx1*^{cre} with MADM alleles to map neuronal and glial lineages in the forebrain as before[17,18,20,34]. MADM allows simultaneous labeling and genetic manipulation in clones of somatic cells in isolated and sparse populations of progenitors[35] and has been successfully employed in genotype-phenotype studies on neurogenesis and gliogenesis[21,36]. MADM-labeled neurons and glia can possess distinct genotypes, which was tracked by permanent labeling of two mitotically-derived daughter cells with two distinct and nonoverlapping fluorescent markers: enhanced green fluorescent protein (eGFP) and/or tdTomato (tdT; Figure 1A). MADM cells in forebrain samples can be counted and classified into six groups based on morphology and color: red (tdT), green (eGFP) and yellow (both reporters co-expressed) neurons and glia (Figure 1B).

The notation *Emx:MADM:+/+* represents wildtype (*WT*) mice, whereas *Emx:MADM:F/F* indicates mice with complete deletion of *Egfr* in the dorsal forebrain (cerebral cortices and hippocampal formation). Both groups of mice have sparse MADM-labeled populations of cells in the dorsal forebrain with identical genotypes (i.e., *WT* or *Egfr-null*, respectively). The *Emx:MADM:F/+* mice are unique in that the vast majority of cells in the dorsal forebrain are heterozygous for *Egfr* deletion, but spare populations of MADM neurons and glia are genetically guided to have three distinct genotypes reported by red, green, or both (yellow) MADM colors: *WT* - red MADM, homozygous *Egfr-null* - green MADM, and heterozygous *Egfr-null* yellow MADM (for details on the genetics of these mice see Zhang et a.l[32]). In other words, the red MADM cells in the dorsal *Emx:MADM:F/+* forebrain express EGFR in excess to their bulk environment, and the green MADM cells lack EGFR expression. The yellow MADM cells are genotypically identical to the majority non-MADM background cells which are all heterozygous for the *Egfr-null* allele. Therefore, a single mouse (*Emx:MADM:F/+*) provided for both overexpression and knockout experimental conditions in the same tissue, which is an advantage of MADM genetics for studying gene-microenvironment interactions.

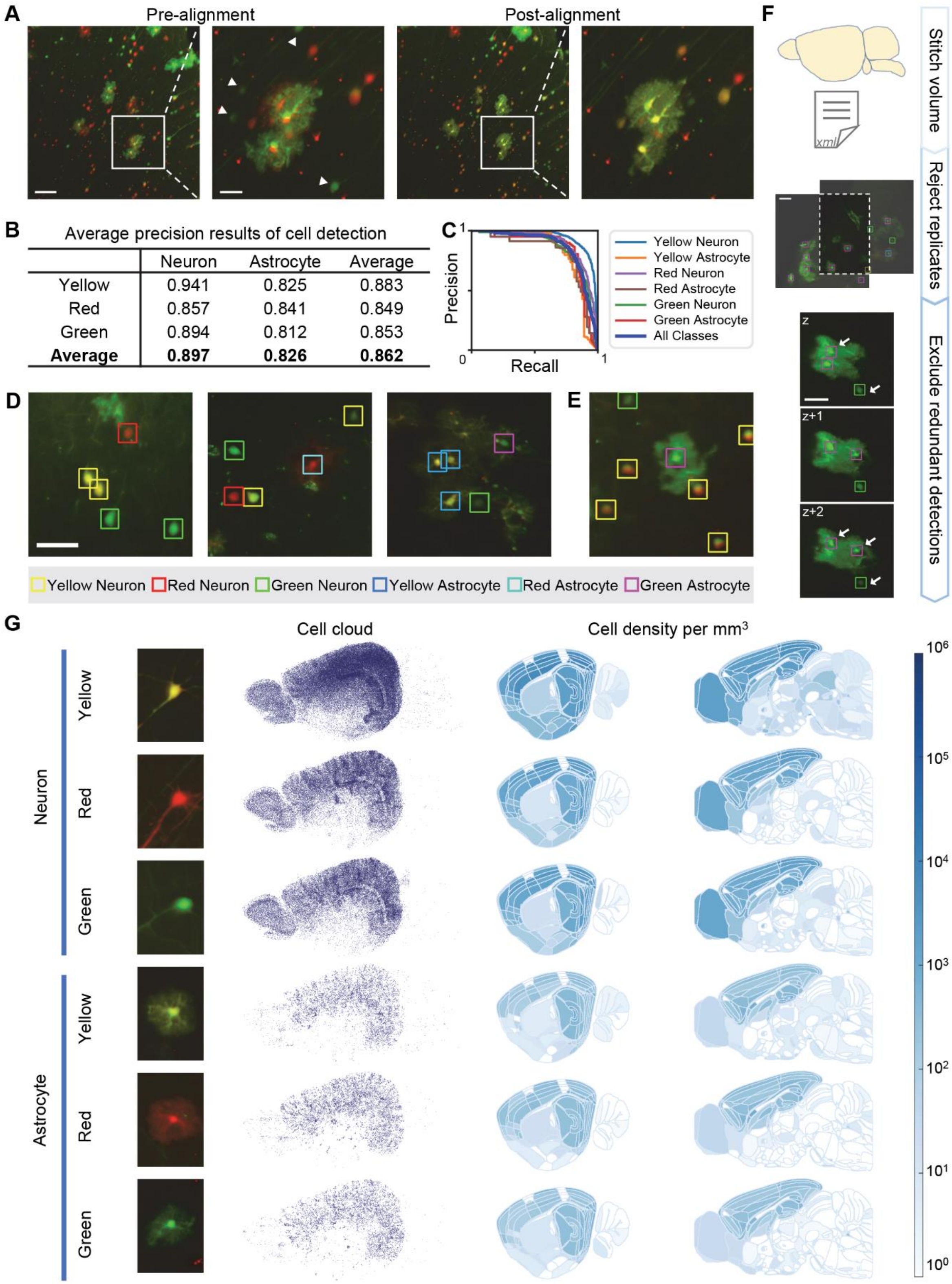
Increased understanding of the complexities in MADM data in the mouse forebrain inspired the need for classification of MADM labeled cells based on morphology and color. Therefore, we designed a workflow that spanned tissue clearing to quantitative analysis (Figure 1C). *Emx:MADM* mouse brains were split into hemispheres, cleared using the iDISCO+ protocol[5] and imaged by a custom-built LSFM[37–39]. The LSFM was equipped with an autofocus algorithm to correct the shift between the light sheet illumination beam and the detection focal plane, and therefore, our images were crisp and focused throughout acquisition. Before imaging, illumination beams of different wavelengths were aligned parallel to ensure accurate data processing (see later). Imaging with subcellular resolution, the acquired two-channel datasets consisted of ~ 400 GB of data per hemisphere. Channel alignment was achieved using NuMorph[15] that locally evaluated the translation between the color channels, followed by a RetinaNet model trained to detect six classes of labeled cells simultaneously[29,30]. The aligned tiles and detected cells were then stitched according to TeraStitcher[40] and registered to the ABA by ClearMap[5] for regional analysis in the forebrain. By mixing and matching codes from different packages we were able to build on their relative advantages (Figure S1). Taking advantage of the abovementioned tools, we were able to analyze a single *Emx:MADM* hemisphere in four days using a regular desktop computer and employing the proposed pipeline. For completeness, glia labeled using *Emx:MADM* forebrains consisted of oligodendrocytes and astrocytes. However, the morphology and size of mature astrocytes was highly distinguishable, whereas oligodendrocytes were small and difficult to detect, and their numbers were small compared to neurons and astrocytes[32]. Thus, we trained our platform to distinguish between neurons and astrocytes.

7.2 COMBINe is highly suitable for automated detection, stitching and registration of MADM cells in the P30 cleared forebrain

RetinaNet was adapted to detect six classes of MADM cells (yellow, red, and green: neurons and astrocytes). RetinaNet is a state-of-the-art object detection network that includes a backbone of a Feature Pyramid Network on top of a ResNet to generate a multi-scale feature pyramid, with two subnetworks to regress and classify anchor boxes[29]. The output of RetinaNet is a list of bounding box coordinates, classification labels, and confidence scores for the detected objects. We selected the RetinaNet model because of its superior performance in dense object detection of real-world images[41] and in cell detection of fluorescence microscopy images[31].

Since local translation was sufficient to register the color channels under parallel beam condition, cell detection was performed after color registration (Figure 2A; see next section). Cell detection was implemented in 2D instead of 3D, due to the difficulty of labeling and training, as well as the variance in resolution along different axes in 3D images. We explored the training regime of the network and applied it to fluorescence images of MADM brain sections in a pervious study[30]. We again employed the data augmentation of color swap and saturation to balance the color classes and utilized transfer learning to train the network on a pretrained backbone. After training, the model which performed best on validation data was selected as the inference model. The inference model reached an average precision of 0.862 across six classes on unseen test data (Figure 2B). An average precision of 0.897 was achieved for detection of neurons, while an average precision of 0.826 was achieved for detection of astrocytes. Given a confidence threshold of 0.5, the mean values of precision, recall, and F-score across six classes on test data were 0.848, 0.771, and 0.806 respectively (Table S1-3). The adjusted precision-recall curves used to measure average precision across six classes are shown in Figure 2C with several examples of the model predications shown in Figure 2D. Color classification was robust even in cases with minor misalignment between the two-color channels (Figure 2E). Examples of manual annotation versus automated prediction on a field of view are shown in Figure S2.

Figure 2. Automated detection, stitching and registration of MADM cells. (A) Images pre- and post- channel alignment from parallel illuminations (left panel; scale bar, 100 µm). Cells in the wrong z-plane (arrowheads) disappear following alignment. Boxed areas are zoomed in images (scale bar, 30 µm). **(B)** An overall average precision of 0.86 was achieved across the six cell classes using RetinaNet. **(C)** Precision-recall curves. **(D-E)** Representative images in which the program made predictions with accurate (D) and inaccurate (E) color registrations (scale bar, 50 µm). See also Figure S2. **(F)** Stitching and merging of detected cells in 3D for each tile were according to TeraStitcher parameters. Replicates in overlapped regions across tiles were rejected, and redundant detections across different z-planes were excluded according to signal intensities. Scale bar, 50 µm. **(G)** Detected cell clouds and corresponding regional densities after registration to the Allen Brain Atlas (ABA), in a single hemisphere of an *Emx:MADM:+/+* forebrain.


A
Pre-alignment
Post-alignment
B
Average precision results of cell detection
Neuron Astrocyte Average
Yellow 0.941 0.825 0.883
Red 0.857 0.841 0.849
Green 0.894 0.812 0.853
Average 0.897 0.826 0.862
C
Precision
Recall
0
1
1
Yellow Neuron
Yellow Astrocyte
Red Neuron
Red Astrocyte
Green Neuron
Green Astrocyte
All Classes
D
E
Yellow Neuron
Red Neuron
Green Neuron
Yellow Astrocyte
Red Astrocyte
Green Astrocyte
F
Stitch volume
xml
Reject replicates
Exclude redundant detections
z
z+1
z+2
G
Cell cloud
Cell density per mm3
Neuron
Yellow
Red
Green
Astrocyte
Yellow
Red
Green
10^6
10^5
10^4
10^3
10^2
10^1
10^0

After detecting the cells in raw tiles, the displacement parameters between individual tiles were estimated using TeraStitcher[40], but the high-resolution volume was not stitched. A custom script using the displacement parameters was created for excluding duplicate detections in overlapping regions between adjacent tiles and removing redundant detections across different z-planes (Figure 2F; see Methods). Analysis of a cleared MADM hemisphere was accomplished in four days using a standard computer with generated datasets containing approximately 100,000 images per sample (more than 1.6 million image patches for cell detection). More than 400,000 MADM cells were detected in each cleared hemisphere. A low-resolution stitched volume of the hemisphere was later registered to brain regions annotated by ABA using ClearMap. Detected MADM cells and corresponding regional density maps of a representative *Emx:MADM:+/+* sample are visualized in Figure 2G using a custom script. Nearly all *Emx:MADM:+/+* cells were in the dorsal forebrain consisting of the olfactory bulb, the cortex, and the hippocampus, which was consistent with previous observation on 2D slices[19]. As a quality check for our color registration process, the yellow, red, and green neurons portion of the total population was approximately 50%, 25% and 25% respectively, which was in line with previous estimates[35].

In summary COMBINe can detect and classify cells in an entire MADM mouse brain hemisphere. The raw datasets were acquired using our custom adaptive LSFM, and since commercial LSFMs are not adaptive, and do not change their properties on the fly, we have tested the effect of two realistic acquisition scenarios on the performance of the RetinaNet model in the next section: 1) when the cells are slightly out-of-focus, 2) when the illumination beams (one per color) are not perfectly parallel.

7.3 RetinaNet accurately detects out-of-focus cells

LSFM images can seem out-of-focus when the objective focal plane does not perfectly overlap with the illumination beam[42]. Therefore, we tested the performance of our network in images containing blurry cells using the same tiles that were acquired in- and out-of-focus. To our surprise, the RetinaNet model performed well in detection and color classification of cells although degradation in image quality was observed in the out-of-focus images. This finding illustrated the robustness of the RetinaNet model, and its ability to generalize well in less-than-optimal conditions (Figure 3A). We synthetically defocused the test data according to[43] to compare the model performance with and without defocus and found negligible decrease in performance within 5 µm from the focal plane (Figure S3). We chose to defocus our images synthetically, because 1) it allows for direct evaluation of model performance in blurry images; 2) labeling real cells in out-of-focus LSFM datasets is ambiguous, as some cells may seem sharp but very dim, and their accurate z position is difficult to derive; 3) it will enable future use of this approach to augment existing datasets. It is important to note that while training the RetinaNet model, cells from the edges of the field-of-view that exhibited lower contrast in comparison to the center of the field-of-view were included. This phenomenon of fluctuations in image quality as a function of position within the field-of-view was an inherent property of using a gaussian illumination beam. Consequently, the network was trained with variability in image contrast, which may explain the robustness of the network in correctly detecting out-of-focus cells (Figure 3A parallel beam scenario).

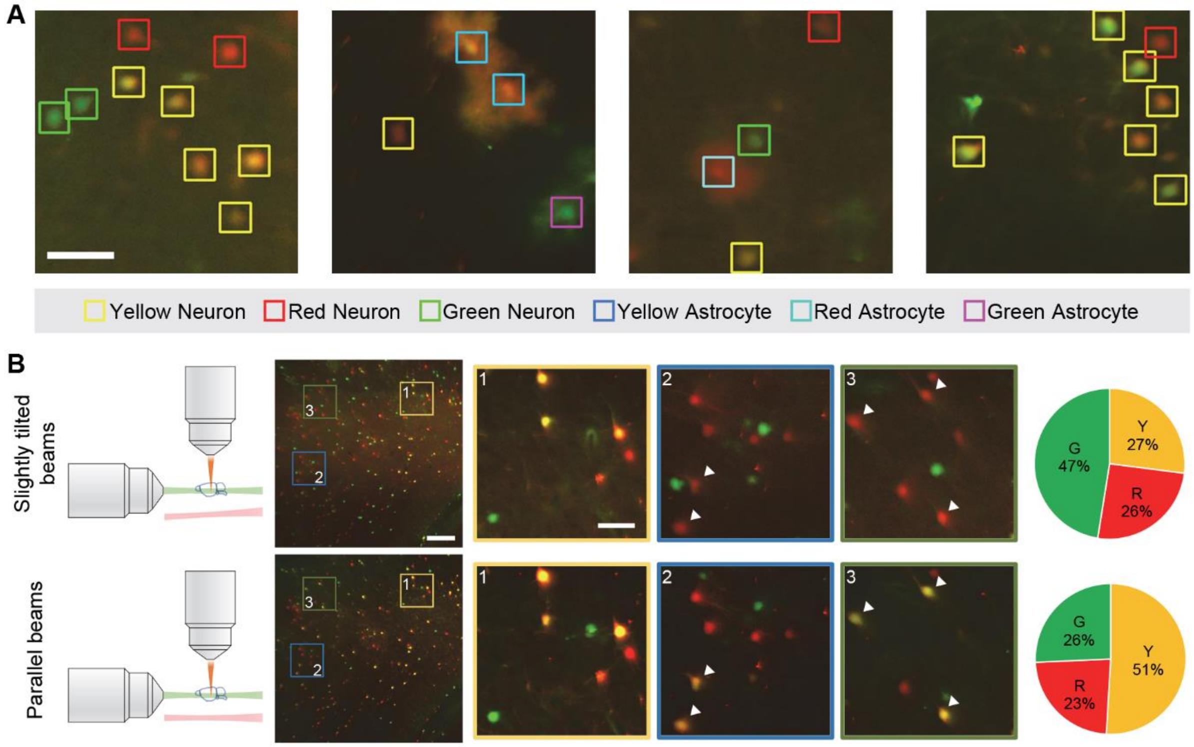


Figure 3. RetinaNet detection is robust for out-of-focus cells. (A) Representative predictions in images of out-of-focus cells (scale bar, 50 μm). See also Figure S3. **(B)** Illustrations of LSFM geometry for tilted (upper) and parallel (bottom) illuminations. Images from the same field-of-view (scale bar, 200 μm) are shown with three zoomed in areas (scale bar, 25 μm). Unparallel beams during imaging result in misalignment between the two channels (e.g., the two-color channels are appropriately overlapped in zoomed region 1 but misaligned in regions 2 and 3 as marked by arrowheads). Erroneous color classification ensues, resulting in fewer yellow cells detected (pie charts).

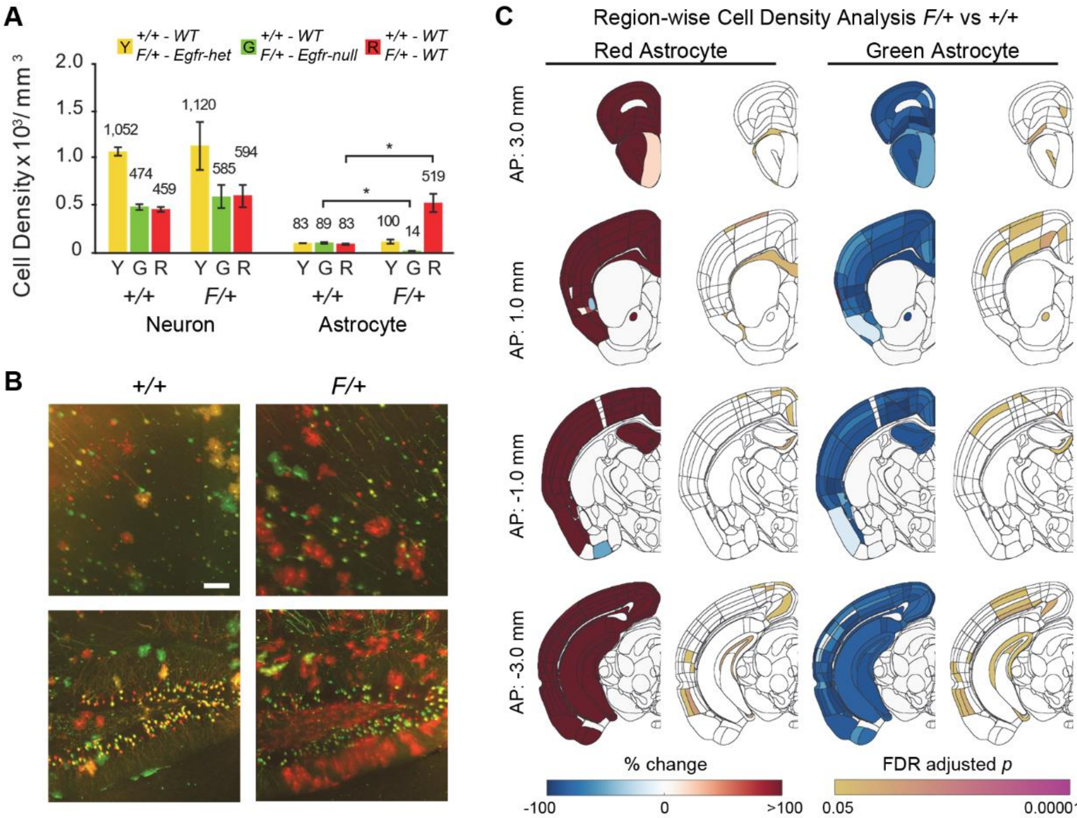
We next tested our pipeline when the illumination beams of different color channels propagated through slightly different light paths during LSFM imaging. This can occur when the microscope setup is out of alignment due to angular drift of the scanning galvo regarding to the room temperature, shift of the laser beam after long period of imaging (e.g., air bubbles) and more. This situation resulted in only partial overlap of the color channels i.e., misalignments in some regions even after registration of the color channels using NuMorph. Importantly, misalignment of the color channels influenced the subsequent quantitative analyses (see pie charts in Figure 3B). In wildtype MADM mice, we expected a color ratio of 50% yellow, 25% red, and 25% green neurons, and we have used this distribution to test the performance of our color registration. This approximate ratio was achieved when the illumination beams were parallel to each other, and the different color channels were registered using local x-y-z translation. However, with slightly tilted illumination beams this ratio was greatly distorted. We observed fewer yellow cells and more green cells.

Together these findings suggest that the pipeline and the RetinaNet model were robust and highly effective for detection and color classification, even in cases containing unfocused images, or slight misalignments between the color channels, as long as the light sheet illumination beams were parallel to each other.

7.4 COMBINe faithfully captures and detects elevation of red (wildtype) astrocytes and depletion of green (Egfr-null) astrocytes in the Emx:MADM:F/+ forebrain

To test the capacity and accuracy of COMBINe we applied it to known phenotypes that we had already captured in *Emx:MADM:F/+* forebrains where neuronal production is largely preserved while gliogenesis is disrupted during perinatal development[32]. As described earlier, red and green cells in *Emx:MADM:F/+* forebrain correspond to *WT* and *Egfr-null* genotypes, respectively. COMBINe confirmed this effect in that both red neurons and green neurons were present and relatively similar in density, while a significant increase in *WT* red MADM astrocytes and a severe absence of green *Egfr-null* MADM astrocytes was observable throughout the labeled regions of the forebrain (Figure 4A and 4B). In addition, no significant differences in total volume of the hemisphere, or regional volumes were observed, suggesting that heterozygous bulk deletion of *Egfr*, or the sparse MADM phenotypes do not cause changes in overall forebrain architecture. Figure 4C shows further region-wise comparison of red and green astrocyte densities between *Emx:MADM:+/+* and *Emx:MADM:F/+* brain hemispheres. Visualized in the left columns are percentage changes of regional cell densities while adjusted *p* values obtained from unpaired t-tests are displayed on the right (see also Figure S4). Among 1326 ABA-annotated regions, 31 regions showed significant elevation of *WT* red astrocytes in the *Emx:MADM:F/+* forebrain (Table S4).

Figure 4. Elevated density of red MADM astrocytes and absence of green MADM astrocytes is faithfully captured by COMBINe in *F/+* cortices. (A) Cell density comparison between *Emx:MADM:+/+* and *Emx:MADM:F/+* mouse brain hemispheres (mean ± SD, number of animals = 3; *, $p < 0.05$). In the *Emx:MADM:F/+* forebrain, red and green cells correspond to *Egfr-null* and *wildtype* genotypes, respectively. While both red and green neurons were present, a substantial increase in wild type red astrocytes was revealed by COMBINe. **(B)** Representative images of the cortex and the hippocampus. Scale bar, 100 µm. **(C)** Region-wise analysis of average red and green astrocyte densities between *Emx:MADM:+/+* and *Emx:MADM:F/+* datasets. Left panels: percentage change in average cell densities of *Emx:MADM:F/+* brain hemispheres compared to *Emx:MADM:+/+* brain hemispheres. Right panels: adjusted *p*-values (number of animals = 3). See also Figure S4 for analysis of other cell types. AP – anteroposterior distance taken from the bregma, FDR – false discovery rate.



A
Y +/+ - WT F/+ - Egfr-het
G +/+ - WT F/+ - Egfr-null
R +/+ - WT F/+ - WT
Cell Density x 10³/ mm³
2.0
1.5
1.0
0.5
0
1,052
474
459
1,120
585
594
83 89 83
100 14
519
Y G R
+/+
F/+
Neuron
Astrocyte
B
+/+
F/+
C
Region-wise Cell Density Analysis F/+ vs +/+
Red Astrocyte
Green Astrocyte
AP: 3.0 mm
AP: 1.0 mm
AP: -1.0 mm
AP: -3.0 mm
% change
-100
0
>100
FDR adjusted p
0.05
0.00001

7.5 Using hierarchical analysis to discover major trends in regional densities of astrocytes

While statistical analysis of all ABA annotated forebrain regions (~1000) was useful, the average size of each brain region was relatively small. As such, the accuracy of registration to the ABA reference atlas is crucial to identify significantly altered densities of cell types, wherein small errors in the registration process can potentially mask significant differences. Therefore, we sought to merge regions based on anatomical similarities using the hierarchical structure of annotation in the ABA and implementing hierarchical analysis of region-wise statistics (Figure 5A, e.g., three-level hierarchical annotation of the somatomotor areas). Comparison of forebrain regions at multiple hierarchies revealed significant differences in red MADM astrocyte densities (Figure 5B; see also Table S5 for regions showing significant differences). Among the high-hierarchy regions, the corpus callosum showed significant increase of red astrocyte density in *Emx:MADM:F/+* brain hemispheres compared to the control group (Figure 5C and 5D), which was consistent with manual observation of the acquired images.

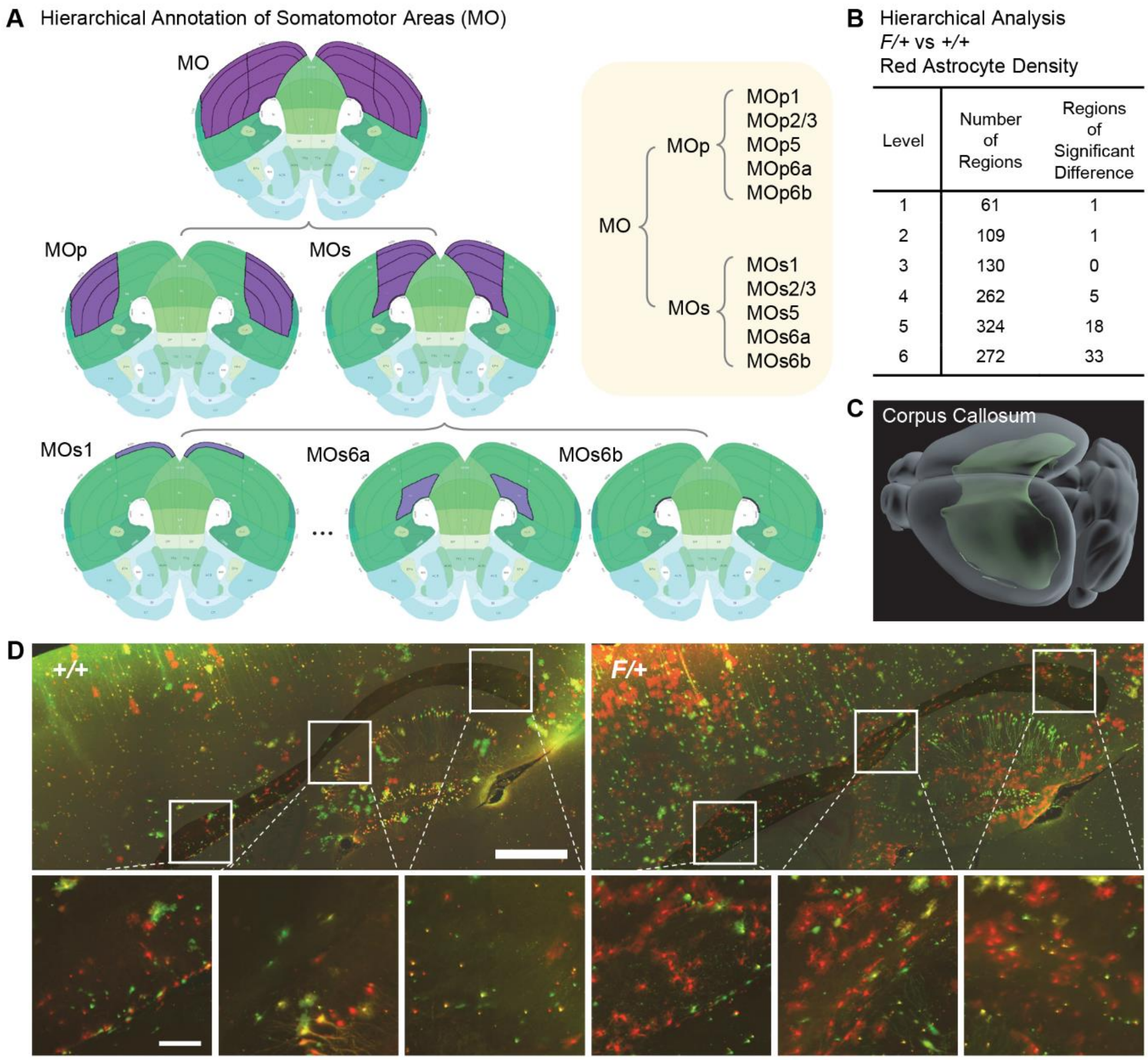


Figure 5. Hierarchical analysis reveals elevation in red MADM astrocytes in the *Emx:MADM:F/+* corpus callosum. (A) Example of hierarchical annotations of somatomotor (MO) cortical areas according to the ABA. Mop – primary motor area. MOs – secondary motor area. **(B)** Results of hierarchical analysis revealed that the corpus callosum contains significantly elevated densities of red MADM astrocytes at hierarchy level 1. **(C)** 3D structure of the corpus callosum provided by ABA. **(D)** Representative images of the corpus callosum (scale bar, 500 μm) and zoomed in regions (scale bar, 100 μm).

7.6 Variance in astrocyte distribution between rostral and caudal *Emx:MADM:F/F* cortices is quantitatively captured using COMBINe

Analysis of neurons and astrocytes in the *Emx:MADM:F/F* forebrain hemispheres using COMBINe confirmed a phenotype analyzed in our recent publication in rostral-caudal differences in EGFR-dependent and EGFR-independent astrocyte production[32]. It is important to be reminded that all cells in the *Emx:MADM:F/F* forebrain are *Egfr-null* regardless of their MADM color. Figure 6A shows the average region-wise cell density maps onto the *Emx:MADM:F/F* isocortex as defined by the ABA. Hence, we compared the average astrocyte densities of selected regions and found that there was a significant difference in astrocyte distribution between rostral and caudal regions (Figure 6B). While astrocyte densities of control *Emx:MADM:+/+* and *Emx:MADM:F/F* brain hemispheres appeared similar in caudal regions, the rostral regions of *Emx:MADM:F/F* cortices contained few, if any, astrocytes (Figure 6C). As expected, neurons that in general do not express EGFR showed similar distributions in the *Emx:MADM:F/F* and *Emx:MADM:+/+* (Figure S5).

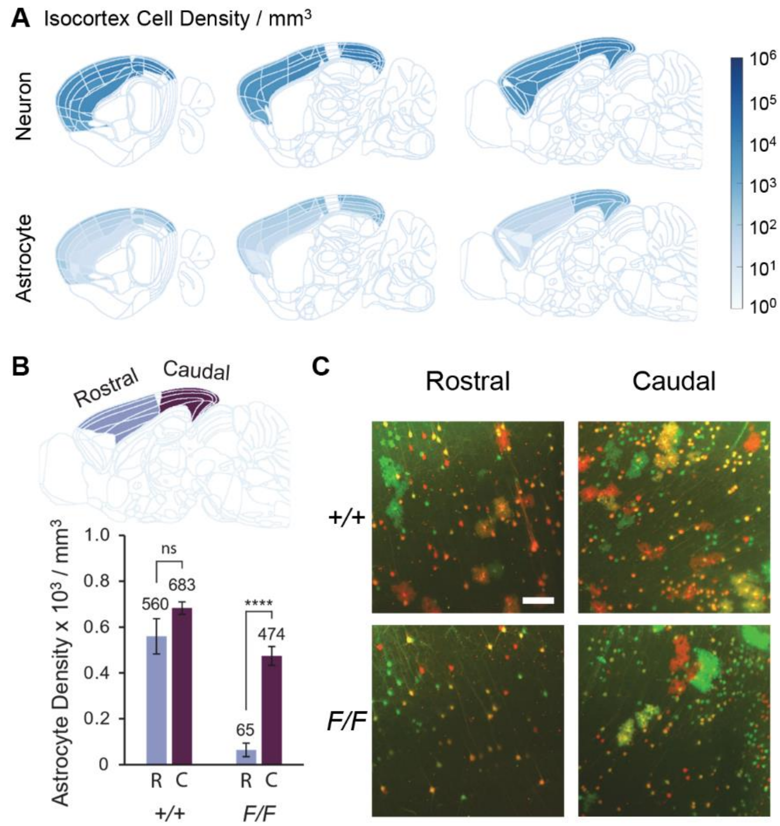


Figure 6. Variance in astrocyte distribution between rostral and caudal cortices is quantitatively captured using COMBINe. (A) Average regional cell density maps of *Emx:MADM:F/F* brain hemispheres (number of animals = 3). **(B)** Significant difference in MADM-labeled astrocyte densities was observed between rostral and caudal regions in *Emx:MADM:F/F* brain hemispheres (bar chart: mean ± SD, number of animals = 3; ****, $p <$ 0.0001). See also Figure S5 for analysis of neuron densities. **(C)** Representative images of rostral and caudal areas (scale bar, 100 μm). Almost no MADM-labeled astrocytes are observable in rostral areas of *Emx:MADM:F/F* brain hemispheres.

To summarize, we automatically calculated and compared the densities of sparse neurons and astrocytes in distinct *Emx1^cre^* induced MADM brain hemispheres with and without conditional deletion of *Egfr*. Region-wise analysis after registering the samples to ABA guided us to devise a hierarchical analysis method to characterize the variances observed in different cell populations among distinct genotypes. The revealed patterns matched laborious manual counting results which validated the COMBINe approach with high confidence.

8 Discussion

In this manuscript, we present COMBINe, an automated framework that can be applied to 3D imaging datasets that are captured with multiple color channels. To demonstrate the applicability of COMBINe, we imaged tissue cleared MADM-labeled mouse forebrains with cellular resolution. We have tested the influence of out-of-focus artifacts on the cell classification accuracy and show that if the illumination beams are parallel the RetinaNet performs adequately. COMBINe uses minimal storage space, as the datasets are not repeatedly saved per pipeline stage (e.g., after stitching and registration with the Allen Brain Atlas). We accomplish that by applying a bookkeeping strategy that uses the calculated transformation matrices in the stitching and registration parts to move the detected cells' coordinates from one representation to the other. Consequently, only the raw tiles and a low-resolution stitched volume are required to complete the analysis. We also introduce a complementary method to perform the statistical analysis of our results, by comparing the cell densities on the hierarchical structure of the brain, provided by the ABA annotation. This method allows to average the cell densities across larger brain areas, that are morphologically and functionally connected, and to minimize the penalty for multiple comparisons. COMBINe can be adapted for other neuroscience studies that require quantitation of elements in the mouse brain using tissue clearing and immunofluorescence labeling. By enriching the framework components (e.g., adding features of segmentation and more), COMBINe can be applied to diverse neuroscience questions in need of analyzing 3D mouse brain datasets.

To test the RetinaNet model detection and classification success we have used the average precision measure, which is a widely used evaluation metric for object detection tasks[44]. The inference model (specific RetinaNet model) that is employed in the workflow achieved an average precision of 0.86 across six classes, showing its efficacy in cell detection and classification. However, when applied to large datasets, sample-wise quality control and common sense is necessary when utilizing learning-based methods. For example, in this study, the color ratio of neurons is referred to as an internal control to assess the cell detection process. Additionally, in the control brains, assessment that the color ratio is maintained across all tiles is used to assess outlier tiles.

Although we report an average precision of only 86%, such evaluation is performed based on human-annotated ground truth. Since there is a considerable variability even between human annotators, there are no perfect models for automated analysis. Hence, it is crucial to keep the analysis process consistent and eliminate potential biases.

Most of the data processing procedures in the current study were performed in 2D given the large size of the datasets and the differences in resolution along axes. In comparison with 3D processing, 2D processing has been intensely investigated and is applicable to diverse image data. Also, the labeling and manipulation of data in 2D is easier and more intuitive than in 3D. However, there are situations where 3D processing is favored. For example, when segmenting and tracing nerves, the structural information is mostly maintained in 3D. To summarize, processing in 2D or 3D should be carefully chosen given specific needs in analysis.

All in all, performance of COMBINe in analysis of *Emx1*^{cre} MADM forebrains is in line with results obtained by manual counting in several ways: 1) restricted distribution of MADM cells, 2) quantitative differences between neurons and astrocytes, and 3) the distribution of different cell colors across the brain. We did not observe volumetric changes when the experimental groups were compared with the control, confirming that sparse alterations using MADM which is estimated to occur in 1:1000 cells in the brain[34,35], has undetectable effects on overall tissue development and homeostasis.

From the biological perspective, COMBINe quantitatively and unbiasedly confirms two important phenotypes in the three *Egfr* genetic backgrounds described here. The findings from *Emx:MADM:F/+* forebrains suggest that when faced with a largely *Egfr-heterozygous* forebrain background, sparse populations of red *wildtype* MADM cells may overexpand to generate a larger number of astrocytes than their normal capacity during perinatal development when gliogenesis occurs. In contrast, *Egfr-null* MADM cells fail to generate astrocytes in the rostral part of the isocortex. We found similar results in comprehensive study of the role of EGFR in forebrain gliogenesis[32], and a clonal study using tamoxifen-inducible *Nes-creER:MADM:F/+* alleles to label gliogenic progeny at clonal densities[21]. Since the *WT* glial progenitors are essentially overexpressing EGFR relative to their surrounding cells in the *Emx:MADM:F/+* background, these results constitute a gain-of-function scenario that explains the strong effect on astrocyte production. An unexpected result in our current study was the region-restricted presence of glia in *Emx:MADM:F/F* forebrains, since the *F/+* findings suggested that no astrocytes should be produced in the *homozygous Egfr-null* background. However, the astrocytes in this case were largely in the caudal, but not rostral forebrain regions. This suggests the possibility of an EGFR-independent gliogenic progenitor pool in the caudal aspect of the forebrain, appears capable of compensating for the loss of EGFR-dependent glia which are critical for astrocyte production in rostral forebrain regions[32]. Interestingly, the EGFR-dependent astrocytes appear to occupy neocortical and possibly other neo-forebrain areas, whereas the EGFR-independent populations which are present in the *Emx:MADM:F/F* appear to seed evolutionary older paleocortical and hippocampal areas. The precise identity and anatomical domain of the EGFR-independent glia producing progenitors remains to be determined.

9 Limitation of Study

Currently, deep learning is still considered a black box since the relationships between input data and the neural networks that 'learn' patterns are not clearly explainable, which makes it difficult to evaluate the errors and failures and improve the model performance[45]. Moreover, as deep learning is a data-driven method, additional training is required before applying it to different sample types. Furthermore, the packages utilized/modified in this study were built on different operating systems (e.g., Linux and Windows) and modifying them to run in a different operating system is challenging and inconvenient. In addition, performance of the brain registration approach adopted in this pipeline should be in line with previous publication that tested ClearMap/Elastix relative to other registration approaches. However, one should be aware that the accuracy of the registration to the ABA drops if the sample phenotype includes anatomical defects. This issue highlights the need for the presented hierarchical approach to compare regional changes in the brain, but also the continued need for validation by the human expert. Last, some of the steps presented in COMBINe require manual check-ins to ensure proper functioning. For example, it is highly recommended to manually inspect a few images in each tile after channel alignment to ensure accuracy, and the cell detection model needs to be validated every time when analyzing a new batch of datasets.

10 Methods

10.1 Resource availability

10.1.1 Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Alon Greenbaum (greenbaum@ncsu.edu).

10.1.2 Materials availability

This study did not generate new unique reagents.

10.1.3 Data and code availability

- An example mouse brain dataset has been deposited at Dryad and is publicly available as of the date of publication. The DOI is listed in the key resources table.
- All original code has been deposited at Zenodo and is publicly available as of the date of publication. The DOI is listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

10.2 Experimental model and subject details

Mice were housed in a 12-hour light:dark cycle with ad libitum access to food and water. All procedures were performed under the regulations and approval from Institutional Animal

Care and Use Committee and at North Carolina State University. *Emx1*^{cre}*:MADM11*$^{TG/GT}$*:Egfr*$^{+/+}$, *Egfr* $^{F/+}$ and *Egfr*$^{F/F}$ mice were generated using breeding schemes previously described[21]. Six males and three females were sacrificed at P30. Three brain hemispheres collected from three mice of each genotype were tissue cleared and analyzed. Sex was not considered as a variable in this study.

10.3 Method details

10.3.1 Tissue clearing

The iDISCO+ protocol[5] was used to stain and clear the mouse brain hemispheres. Brain hemispheres were dehydrated at room temperature (RT) using methanol gradients diluted in ddH_2O: 20%, 40%, 60%, 80%, 100%, 100%; 1 h each. Brain hemispheres were then chilled at 4 °C for 2 h and submerged in the mixture of dichloromethane and methanol (v:v=2:1) overnight at RT. After washing twice in 100% methanol at RT, brain hemispheres were chilled at 4 °C for 2 h and decolorized in chilled fresh 5% H_2O_2 in methanol overnight at 4 °C. After bleaching, brain hemispheres were rehydrated with methanol series diluted in ddH_2O (100%, 80%, 60%, 40%, 20%), followed by one wash with PBS and two washes with PBS with 0.02% Triton X-100, 1 h each at RT. Next, brain hemispheres were permeabilized with permeabilization solution (0.2% Triton X-100, 0.3M Glycine, 20% DMSO in PBS) for 2 d at 37 °C and blocked with blocking solution (0.2% Triton X-100, 6% donkey serum, 10% DMSO in PBS) for 2 d at 37 °C. Brain hemispheres were then incubated with chicken anti-GFP (1:200, Aves Labs) and rabbit anti-RFP (1:200, Rockland) in PTwH (0.2% Tween-20, 10 mg/L heparin in PBS) with 5% DMSO (v/v) and 3% donkey serum for 7 d at 37 °C. After primary staining, brain hemispheres were washed 5 times with PTwH, 2 h each until the next day. Next, brain hemispheres were incubated with donkey anti-chicken AlexaFluor 647 (1:200, Jackson ImmunoResearch) and donkey anti-rabbit Cy3 (1:200, Jackson ImmunoResearch) in PTwH with 3% donkey serum for 5 d at 37 °C, followed by 5 washes with PTwH. For optical clearing, brain hemispheres were dehydrated with methanol series diluted in ddH_2O (20%, 40%, 60%, 80%, 100%), 1 h each at RT, and incubated in the mixture of dichloromethane and methanol (v:v=2:1) for 3 h at RT. Brain hemispheres were then submerged in dichloromethane 15 minutes twice at RT. All the incubation steps above were performed with gentle shaking. Brain hemispheres were cleared with dibenzyl ether at RT.

10.3.2 Light-sheet imaging

Cleared mouse brain hemispheres were imaged using a custom-built light sheet fluorescence microscope, of which the setup was outlined in[37–39]. Each brain hemisphere was imaged by illumination of 561 nm and 640 nm respectively, with a voxel size of 0.65 × 0.65 × 10 μm^3. During image acquisition, every 1 mm in z direction, autofocus method would be applied to correct for the drift between the light-sheet plane and the detection focal plane.

10.3.3 Channel alignment

When the illumination beams of different wavelengths are parallel, images of different color channels can be matched through translation in x, y, and z-axes. NuMorph[15] was employed for local channel alignment on each imaged tile (MATLAB). First, displacement in z-planes was determined by maximizing phase correlation between two channels. After adjustment in z, color registration was performed in x- and y-axes.

10.3.4 Cell detection

Both the training data and the validation data contained fluorescence images from different imaging modalities, e.g., slide scanner, confocal fluorescence microscopy, and LSFM[30]. 3361 individual cells (2497 neurons and 864 astrocytes) were labeled in the training data. 1164 individual cells (896 neurons and 268 astrocytes) were labeled in the validation data. Test data was generated from light sheet datasets that were analyzed in this paper. 1664 individual cells (1462 neurons and 202 astrocytes) were labeled in the test data.

A RetinaNet[29] model was trained to detect cells of six classes: yellow, red, green neurons and astrocytes. The repository cloned from the source (https://github.com/fizyr/keras-retinanet) was adapted to this work (https://github.com/yccc12/COMBINe/keras-retinanet). Transfer learning was adopted with a pre-trained ResNet50 being used as the backbone. Data augmentation strategies such as geometrical transforms, color swap and saturation simulation were applied[30]. The initial learning rate was 0.0001 and the batch size was four. Using an Adam optimizer, the model was trained for 50 epochs. The model with the best performance on validation data was selected as the inference model. Average precision measure[44] was adopted to evaluate model performance. The average precision measure is widely used in object detection tasks, and it is very similar to the area under the precision-recall curve. Other measures such as precision, recall, and F-scores[46] were reported based on a confidence threshold of 0.5.

Because stitching was likely to generate artifacts, cell detection was performed per image tile, during which images were cropped into 512 × 512 image patches with an overlap fraction of 0.125 for inference. Detections in overlapped areas were merged based on intersection over union and confidence score.

10.3.5 Stitching

TeraStitcher[40] advanced mode was used to stitch the acquired datasets, with misalignment across tiles being corrected. Parameters output from TeraStitcher were adopted to stitch detected cells of each tile. Duplicate detections in overlapping regions between adjacent tiles were rejected. Redundant detections across different z-planes were merged based on signal intensities when the output bounding boxes between adjacent z-planes were found overlapped, which was determined by an intersection over union above 0.5 between two bounding boxes. Stitched datasets and detected cells were visualized in Imaris 9.5 (Oxford Instruments).

10.3.6 Registration

ClearMap[5] in Linux was utilized to register stitched datasets to Allen Brain Atlas[33] and to map detected cells to annotated brain regions.

10.3.7 Visualization

Cell densities and statistics were visualized in MATLAB using a custom script.

10.4 Quantification and statistical analysis

Regional cell densities between different groups were compared in MATLAB using unpaired t-tests assuming unequal variances. *p*-values were adjusted by the Benjamini–Hochberg procedure[47] to control the false discovery rate (< 0.05) for simultaneous multiple testing. All statistics were reported as mean ± standard deviation (SD).

11 Acknowledgements

The authors would like to thank the Greenbaum lab members for discussions and assistance in revising the manuscript.

Funding: National Institutes of Health (USA) grants to both HTG and AG (R21NS129093 and R56NS117019), and to AG (R21DC020005).

12 Author contributions

Conceptualization: YC, AG, HTG; Methodology: YC, CL; Investigation: YC; Software: YC; Visualization: YC; Resources: XZ, HTG; Supervision: AG, HTG; Writing – original draft: YC; Writing – review & editing: YC, AG, HTG

13 Declaration of interests

The authors declare no competing interests.

14 Inclusion and diversity

We worked to ensure sex balance in the selection of non-human subjects. One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in their field of research or within their geographical location.

REFERENCES

1. Cai, R., Pan, C., Ghasemigharagoz, A., Todorov, M.I., Förstera, B., Zhao, S., Bhatia, H.S., Parra-Damas, A., Mrowka, L., Theodorou, D., et al. (2019). Panoptic imaging of transparent mice reveals whole-body neuronal projections and skull–meninges connections. Nat Neurosci *22*, 317–327. 10.1038/s41593-018-0301-3.
2. Chung, K., and Deisseroth, K. (2013). CLARITY for mapping the nervous system. Nat Methods *10*, 508–513. 10.1038/nmeth.2481.
3. Ertürk, A., Becker, K., Jährling, N., Mauch, C.P., Hojer, C.D., Egen, J.G., Hellal, F., Bradke, F., Sheng, M., and Dodt, H.-U. (2012). Three-dimensional imaging of solvent-cleared organs using 3DISCO. Nat Protoc *7*, 1983–1995. 10.1038/nprot.2012.119.
4. Nudell, V., Wang, Y., Pang, Z., Lal, N.K., Huang, M., Shaabani, N., Kanim, W., Teijaro, J., Maximov, A., and Ye, L. (2022). HYBRiD: hydrogel-reinforced DISCO for clearing mammalian bodies. Nat Methods *19*, 479–485. 10.1038/s41592-022-01427-0.
5. Renier, N., Adams, E.L., Kirst, C., Wu, Z., Azevedo, R., Kohl, J., Autry, A.E., Kadiri, L., Umadevi Venkataraju, K., Zhou, Y., et al. (2016). Mapping of Brain Activity by Automated Volume Analysis of Immediate Early Genes. Cell *165*, 1789–1802. 10.1016/j.cell.2016.05.007.
6. Richardson, D.S., and Lichtman, J.W. (2015). Clarifying Tissue Clearing. Cell *162*, 246–257. 10.1016/j.cell.2015.06.067.
7. Susaki, E.A., Tainaka, K., Perrin, D., Yukinaga, H., Kuno, A., and Ueda, H.R. (2015). Advanced CUBIC protocols for whole-brain and whole-body clearing and imaging. Nat Protoc *10*, 1709–1727. 10.1038/nprot.2015.085.

8. Kubota, S.I., Takahashi, K., Nishida, J., Morishita, Y., Ehata, S., Tainaka, K., Miyazono, K., and Ueda, H.R. (2017). Whole-Body Profiling of Cancer Metastasis with Single-Cell Resolution. Cell Reports *20*, 236–250. 10.1016/j.celrep.2017.06.010.

9. Moatti, A., Li, C., Sivadanam, S., Cai, Y., Ranta, J., Piedrahita, J.A., Cheng, A.G., Ligler, F.S., and Greenbaum, A. (2022). Ontogeny of cellular organization and LGR5 expression in porcine cochlea revealed using tissue clearing and 3D imaging. iScience *25*, 104695. 10.1016/j.isci.2022.104695.

10. Ueda, H.R., Ertürk, A., Chung, K., Gradinaru, V., Chédotal, A., Tomancak, P., and Keller, P.J. (2020). Tissue clearing and its applications in neuroscience. Nat Rev Neurosci *21*, 61–79. 10.1038/s41583-019-0250-1.

11. Zhao, S., Todorov, M.I., Cai, R., -Maskari, R.A., Steinke, H., Kemter, E., Mai, H., Rong, Z., Warmer, M., Stanic, K., et al. (2020). Cellular and Molecular Probing of Intact Human Organs. Cell *180*, 796-812.e19. 10.1016/j.cell.2020.01.030.

12. Frasconi, P., Silvestri, L., Soda, P., Cortini, R., Pavone, F.S., and Iannello, G. (2014). Large-scale automated identification of mouse brain cells in confocal light sheet microscopy images. Bioinformatics *30*, i587–i593. 10.1093/bioinformatics/btu469.

13. Silvestri, L., Müllenbroich, M.C., Costantini, I., Di Giovanna, A.P., Mazzamuto, G., Franceschini, A., Kutra, D., Kreshuk, A., Checcucci, C., Toresano, L.O., et al. (2021). Universal autofocus for quantitative volumetric microscopy of whole mouse brains. Nat Methods *18*, 953–958. 10.1038/s41592-021-01208-1.

14. Matsumoto, K., Mitani, T.T., Horiguchi, S.A., Kaneshiro, J., Murakami, T.C., Mano, T., Fujishima, H., Konno, A., Watanabe, T.M., Hirai, H., et al. (2019). Advanced CUBIC

tissue clearing for whole-organ cell profiling. Nat Protoc *14*, 3506–3537. 10.1038/s41596-019-0240-9.

15. Krupa, O., Fragola, G., Hadden-Ford, E., Mory, J.T., Liu, T., Humphrey, Z., Rees, B.W., Krishnamurthy, A., Snider, W.D., Zylka, M.J., et al. (2021). NuMorph: Tools for cortical cellular phenotyping in tissue-cleared whole-brain images. Cell Reports *37*, 109802. 10.1016/j.celrep.2021.109802.

16. Zhou, B., Kaucka, M., Chagin, A.S., and Newton, P.T. (2019). Clonal Genetic Tracing using the Confetti Mouse to Study Mineralized Tissues. J Vis Exp. 10.3791/60424.

17. Beattie, R., Postiglione, M.P., Burnett, L.E., Laukoter, S., Streicher, C., Pauler, F.M., Xiao, G., Klezovitch, O., Vasioukhin, V., Ghashghaei, T.H., et al. (2017). Mosaic Analysis with Double Markers Reveals Distinct Sequential Functions of Lgl1 in Neural Stem Cells. Neuron *94*, 517-533.e3. 10.1016/j.neuron.2017.04.012.

18. Johnson, C.A., and Ghashghaei, H.T. (2020). Sp2 regulates late neurogenic but not early expansive divisions of neural stem cells underlying population growth in the mouse cortex. Development *147*. 10.1242/dev.186056.

19. Liang, H., Hippenmeyer, S., and Ghashghaei, H.T. (2012). A Nestin-cre transgenic mouse is insufficient for recombination in early embryonic neural progenitors. Biol Open *1*, 1200–1203. 10.1242/bio.20122287.

20. Liang, H., Xiao, G., Yin, H., Hippenmeyer, S., Horowitz, J.M., and Ghashghaei, H.T. (2013). Neural development is dependent on the function of specificity protein 2 in cell cycle progression. Development *140*, 552–561. 10.1242/dev.085621.

21. Zhang, X., Mennicke, C.V., Xiao, G., Beattie, R., Haider, M.A., Hippenmeyer, S., and Ghashghaei, H.T. (2020). Clonal Analysis of Gliogenesis in the Cerebral Cortex Reveals

Stochastic Expansion of Glia and Cell Autonomous Responses to Egfr Dosage. Cells *9*. 10.3390/cells9122662.

22. LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. Nature *521*, 436–444. 10.1038/nature14539.

23. Bayramoglu, N., Kannala, J., and Heikkila, J. (2016). Deep learning for magnification independent breast cancer histopathology image classification. In 2016 23rd International Conference on Pattern Recognition (ICPR) (IEEE), pp. 2440–2445. 10.1109/ICPR.2016.7900002.

24. Hollandi, R., Szkalisity, A., Toth, T., Tasnadi, E., Molnar, C., Mathe, B., Grexa, I., Molnar, J., Balind, A., Gorbe, M., et al. (2020). nucleAIzer: A Parameter-free Deep Learning Framework for Nucleus Segmentation Using Image Style Transfer. Cell Systems *10*, 453-458.e6. 10.1016/j.cels.2020.04.003.

25. Sità, L., Brondi, M., Lagomarsino de Leon Roig, P., Curreli, S., Panniello, M., Vecchia, D., and Fellin, T. (2022). A deep-learning approach for online cell identification and trace extraction in functional two-photon calcium imaging. Nat Commun *13*, 1529. 10.1038/s41467-022-29180-0.

26. Pan, C., Schoppe, O., Parra-Damas, A., Cai, R., Todorov, M.I., Gondi, G., von Neubeck, B., Böğürcü-Seidel, N., Seidel, S., Sleiman, K., et al. (2019). Deep Learning Reveals Cancer Metastasis and Therapeutic Antibody Targeting in the Entire Body. Cell *179*, 1661-1676.e19. 10.1016/j.cell.2019.11.013.

27. Todorov, M.I., Paetzold, J.C., Schoppe, O., Tetteh, G., Shit, S., Efremov, V., Todorov-Völgyi, K., Düring, M., Dichgans, M., Piraud, M., et al. (2020). Machine learning analysis of whole mouse brain vasculature. Nat Methods *17*, 442–449. 10.1038/s41592-020-0792-1.

28. Geng, J., Zhang, X., Prabhu, S., Shahoei, S.H., Nelson, E.R., Swanson, K.S., Anastasio, M.A., and Smith, A.M. (2021). 3D microscopy and deep learning reveal the heterogeneity of crown-like structure microenvironments in intact adipose tissue. Science Advances *7*, eabe2480. 10.1126/sciadv.abe2480.

29. Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2018). Focal Loss for Dense Object Detection. Preprint at arXiv, 10.48550/arXiv.1708.02002 10.48550/arXiv.1708.02002.

30. Cai, Y., Zhang, X., Kovalsky, S.Z., Ghashghaei, H.T., and Greenbaum, A. (2021). Detection and classification of neurons and glial cells in the MADM mouse brain using RetinaNet. PLOS ONE *16*, e0257426. 10.1371/journal.pone.0257426.

31. Waithe, D., Brown, J.M., Reglinski, K., Diez-Sevilla, I., Roberts, D., and Eggeling, C. (2020). Object detection networks and augmented reality for cellular detection in fluorescence microscopy. Journal of Cell Biology *219*, e201903166. 10.1083/jcb.201903166.

32. Zhang, X., Xiao, G., Johnson, C., Cai, Y., Horowitz, Z.K., Mennicke, C., Coffey, R., Haider, M., Threadgill, D., Eliscu, R., et al. (2023). Bulk and mosaic deletions of Egfr reveal regionally defined gliogenesis in the developing mouse forebrain. iScience *26*, 106242. 10.1016/j.isci.2023.106242.

33. Wang, Q., Ding, S.-L., Li, Y., Royall, J., Feng, D., Lesnar, P., Graddis, N., Naeemi, M., Facer, B., Ho, A., et al. (2020). The Allen Mouse Brain Common Coordinate Framework: A 3D Reference Atlas. Cell *181*, 936-953.e20. 10.1016/j.cell.2020.04.007.

34. Hippenmeyer, S., Youn, Y.H., Moon, H.M., Miyamichi, K., Zong, H., Wynshaw-Boris, A., and Luo, L. (2010). Genetic mosaic dissection of Lis1 and Ndel1 in neuronal migration. Neuron *68*, 695–709. 10.1016/j.neuron.2010.09.027.

35. Zong, H., Espinosa, J.S., Su, H.H., Muzumdar, M.D., and Luo, L. (2005). Mosaic Analysis with Double Markers in Mice. Cell *121*, 479–492. 10.1016/j.cell.2005.02.012.

36. Laukoter, S., Beattie, R., Pauler, F.M., Amberg, N., Nakayama, K.I., and Hippenmeyer, S. (2020). Imprinted Cdkn1c genomic locus cell-autonomously promotes cell survival in cerebral cortex development. Nat Commun *11*, 195. 10.1038/s41467-019-14077-2.

37. Li, C., Moatti, A., Zhang, X., Ghashghaei, H.T., and Greenbaum, A. (2021). Deep learning-based autofocus method enhances image quality in light-sheet fluorescence microscopy. Biomed. Opt. Express, BOE *12*, 5214–5226. 10.1364/BOE.427099.

38. Li, C., Rai, M.R., Ghashghaei, H.T., and Greenbaum, A. (2022). Illumination angle correction during image acquisition in light-sheet fluorescence microscopy using deep learning. Biomed. Opt. Express, BOE *13*, 888–901. 10.1364/BOE.447392.

39. Moatti, A., Cai, Y., Li, C., Sattler, T., Edwards, L., Piedrahita, J., Ligler, F.S., and Greenbaum, A. (2020). Three-dimensional imaging of intact porcine cochlea using tissue clearing and custom-built light-sheet microscopy. Biomed. Opt. Express, BOE *11*, 6181–6196. 10.1364/BOE.402991.

40. Bria, A., and Iannello, G. (2012). TeraStitcher - A tool for fast automatic 3D-stitching of teravoxel-sized microscopy images. BMC Bioinformatics *13*, 316. 10.1186/1471-2105-13-316.

41. Zou, Z., Shi, Z., Guo, Y., and Ye, J. (2019). Object Detection in 20 Years: A Survey. arXiv:1905.05055 [cs].

42. Tomer, R., Lovett-Barron, M., Kauvar, I., Andalman, A., Burns, V.M., Sankaran, S., Grosenick, L., Broxton, M., Yang, S., and Deisseroth, K. (2015). SPED Light Sheet Microscopy: Fast Mapping of Biological System Structure and Function. Cell *163*, 1796–1806. 10.1016/j.cell.2015.11.061.

43. Yang, S.J., Berndl, M., Michael Ando, D., Barch, M., Narayanaswamy, A., Christiansen, E., Hoyer, S., Roat, C., Hung, J., Rueden, C.T., et al. (2018). Assessing microscope image focus quality with deep learning. BMC Bioinformatics *19*, 77. 10.1186/s12859-018-2087-4.

44. Everingham, M., Eslami, S.M.A., Van Gool, L., Williams, C.K.I., Winn, J., and Zisserman, A. (2015). The Pascal Visual Object Classes Challenge: A Retrospective. Int J Comput Vis *111*, 98–136. 10.1007/s11263-014-0733-5.

45. Belle, V., and Papantonis, I. (2021). Principles and Practice of Explainable Machine Learning. Frontiers in Big Data *4*.

46. Hossin, M., and Sulaiman, M.N. (2015). A Review on Evaluation Metrics for Data Classification Evaluations. International Journal of Data Mining & Knowledge Management Process *5*, 01–11. 10.5121/ijdkp.2015.5201.

47. Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B (Methodological) *57*, 289–300.

CHAPTER 4: Deep learning-based automated detection of spiral ganglion neurons in intact cochleae

1 Abstract

The first stage of neural processing in the inner ear commences with the activation of the auditory nerve. The spiral ganglion neurons (SGNs) which comprise the nerve are therefore a critical target for cochlear implants and future therapeutic strategies to remediate hearing loss. To shed more light on the SGNs and other cellular populations within the cochlea, 3D information is paramount as the spatial location of the cell within the cochlea reveals information on its functional role, and whether the therapeutic has reached its target cells. To address this gap, tissue clearing, and light sheet fluorescence microscopy were recently applied to intact cochlea for analysis of hair cells and supporting cells. Yet, the SGNs that reside deeper in the inner ear within the modiolus remain relatively understudied. This gap is even more profound in large animal models that mimic humans in terms of size, anatomy, and physiology. Hence, to study the SGNs in large and small animal models we: 1) Introduced collagenase treatment to the cochlea to support uniform antibody staining of SGN within the entire modiolus. 2) Adopted a deep learning object detection model to locate and count SGNs in large 3D datasets. To improve the model accuracy, the physical model of the imaging system was used to simulate defocus aberration that dominates light sheet fluorescence microscopy. Using the above approach, all Type I SGNs in intact gerbil and porcine cochleae were detected and counted in 3D, although the specimens were acquired using two different light sheet microscopes and different staining protocols. We believe that broad adaptation of the method will lead to better understanding of the SGN population and their role during hearing loss.

2 Introduction

Hearing loss is one of the most common sensory deficits, affecting nearly 20% of the global population. The causes for hearing loss include loud noise, aging, drug ototoxicity, chronic ear infections, heredity and prenatal infections [1]. Recently promising clinical trials have been conducted to address monogenic hearing loss using gene therapy [2]. To restore hearing after sensorineural hearing loss, the role of spiral ganglion neurons will continue to be critical. A vast number of therapeutic treatments have been proposed, meanwhile hearing tests (i.e., the gold-standard) are utilized to evaluate the treatment outcome. Although hearing tests provide a thorough and insightful evaluation on the treatment success, details of the treatment effects are lacking for further improvements.

To appropriately evaluate the therapeutic treatment, it is important to study the cochlear structure and function in the inner ear at cellular level, e.g., hair cells, spiral ganglion neurons (SGNs), supporting cells, and more. Additionally, studying the cochlea in 3D is especially critical as the spatial location of the cell within the cochlea reveals information on its functional role in hearing, corresponding to the cochlear frequency map, gradients in ion channels expression and more [3]. However, since the cochlea is entombed in bone, processing the tissue is a technical challenge. The cochlea is typically sliced after extensive decalcification, and in the 2D sections the crucial 3D information is lost.

To profile the cochlea in 3D, tissue clearing techniques which were developed to render intact organs transparent [4–11], have also been adapted to the cochlea. These methods were used to investigate the murine, gerbil, guinea pig, and porcine cochlea [12–18]. To image the large cleared volumes, fast imaging techniques such as light sheet fluorescence microscopy (LSFM) were utilized [13–15,19]. Using tissue clearing, hair cell numbers, and protein distribution [13] within the organ of Corti (OC) have been measured. Yet, the SGNs, which is a clinically important population that relay the auditory signal to the brain stem, have not been thoroughly profiled in entire cochlea 3D space.

Compared to hair cells in the OC, SGNs are embedded in the Rosenthal's canal within the modiolus (Figure S1), which results in a bone inside a bone configuration that is challenging to access. This configuration makes it challenging to extract the entire spiral ganglion (SG) out of a mature cochlea for whole-mount imaging. Moreover, such a structure makes it difficult to uniformly stain SGNs deep inside an intact sample. Only a few studies showed staining and imaging of SGNs in entire rodent cochleae [19,20]. These studies also mainly relied on autofluorescence of the SGNs, as achieving uniform staining through the modiolus is difficult. Given the difficulties with charactering the SGNs in small animal models, labeling and 3D imaging of the SGNs is not achieved in large animal models of which the size and thickness of the bone is at least order of magnitude larger than rodents. Large animal models are important to biodistribution studies considering their cochlea similarity to humans in terms of size and general anatomy and functionally because they share similar hearing ranges to humans especially at lower frequencies.

An additional gap in working with large animals or with 3D datasets in general, is the difficulty in analyzing the resulting large datasets. For instance, 3D datasets of multiple fluorescent channels regularly contain >80 gigabytes and often hundreds of gigabytes of images. Automated and semi-automated methods have been developed for hair cell analysis of both 2D and 3D image datasets [14,16,18,21,22], but no such tools for SGN quantification have been developed. Quantification of SGNs were performed manually [19,20,23,24], which is laborious and time-consuming. Deep learning, which has been widely used to assist and accelerate biomedical data analysis [18,21,22,25–29], has shown its efficiency and efficacy in detection of hair cells in murine cochlea images [21,22] and detection of neurons and astrocytes in mouse brain datasets [28,29]. Hence, deep learning-based methods can also be implemented to automate the processing and analysis of SGN datasets.

Therefore, to study the SGNs in the cochlea of large and small animal models while extracting the precise 3D information provided, and in the future to apply this technology to pre-clinical and clinical studies, we: 1) Optimized the tissue clearing method for large animal models (e.g., porcine cochlea), by using collagenase to improve the antibody penetration to the modiolus, and 2) Adopted a deep learning object detection model (RetinaNet [30]) to locate and count SGNs in the resulting 3D datasets. To increase the generalization of our deep learning model, data from two LSFMs, different sample preparations, animal models and different antibody markers was used. Data augmentation including physical modeling of defocus [31] was also utilized, thus the trained model is robust and can be applied to datasets from different animals or with different markers.

Utilizing the proposed workflow that we named Spiner (Spiral ganglion neuron profiler), Type I SGNs (TuJ1 stained) in the gerbil datasets were automatically located within a few hours and the obtained cell counts are close to manual analysis that took days to achieve per sample. By applying the cochlear frequency map for each species to the acquired 3D cochlea volumes, we could study distribution of SGNs at different frequency locations, which can be compared across samples. We further quantified using Spiner Type I SGNs in the porcine cochleae and reported their numbers. Spiner can be used in varied studies requiring quantification and mapping of SGN survival corresponding to desired cochlear frequency ranges in both 2D and

3D. We hope that broad adaptation of Spiner will lead to better understanding of the overall population of SGNs and their role during hearing loss.

3 Results

3.1 Spiner development – immunostaining, imaging and automated counting of neurons in an intact tissue cleared cochlea

We designed Spiner to immunostain, image, and analyze SGNs in whole-mount cochlea samples (Figure 1A). The cleared gerbil cochleae with immunostaining of TuJ1 were imaged with a commercial LSFM (LaVision UltraMicroscope II Light Sheet Microscope, LaVision BioTec GmbH). For tissue clearing of porcine cochleae, we started with optimizing the BoneClear [4] protocol. Specifically, we added a collagenase treatment procedure that improved antibody penetration into the modiolus, which can be used in cochleae from large animal samples with thicker bones and barriers. The cleared porcine samples were imaged by a custom-built LSFM [14,32,33]. Compared to gerbil cochlea, porcine cochlea is nearly five times larger in each dimension, making it difficult to acquire images with good quality deep in the tissue given the light attenuation and deflection caused by massive bone structure with canals and cavities. Variations in the refractive index between the sample and imaging medium would also change the focal plane of the detection objective, thus blurring the image. To correct for changes in the focal distance at different imaging locations, a manual calibration process was included. Before imaging, focus calibration points were manually adjusted around every 50 z-planes in each channel. During the imaging process, linear interpolation of focus based on calibration parameters was adopted to optimize the resulting image quality. Figure 1B shows representative fluorescence images of tissue cleared gerbil and porcine cochleae with TuJ1 staining, in which individual SGNs can be clearly distinguished. A porcine cochlea dataset of two fluorescence channels contained around one hundred gigabytes.

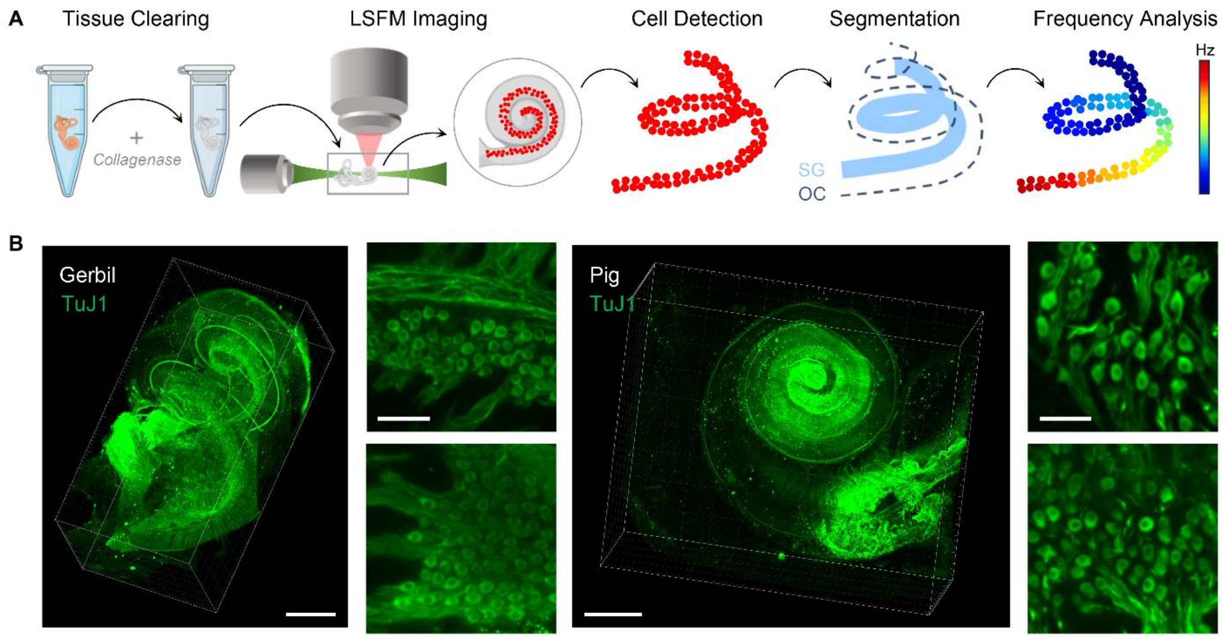


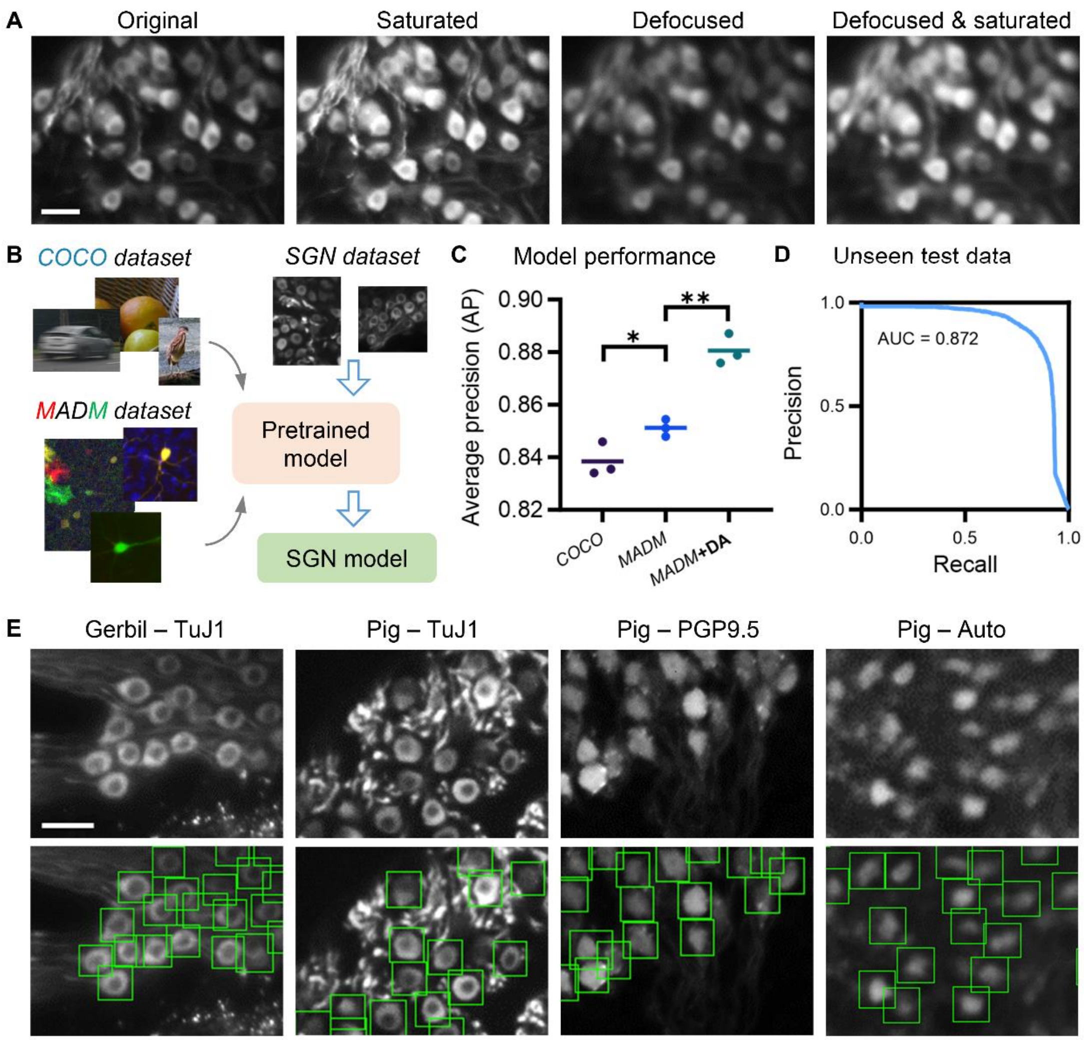
Figure 1. Automated neuron detection workflow for tissue cleared cochlea using Spiner. **(A)** The proposed workflow to locate spiral ganglion neurons in cleared cochlea. **(B)** Left: Maximum intensity projection (scale bar, 500 µm) of fluorescence signals from a gerbil cochlea stained with TuJ1 antibody, with two zoomed in areas at different locations (scale bar, 50 µm). Right: Maximum intensity projection (scale bar, 1 mm) of fluorescence signals from a porcine cochlea stained with TuJ1 antibody, with two zoomed in areas at different locations (scale bar, 50 µm).

To accelerate data analysis, a deep learning model (RetinaNet [30]) was then trained to automatically detect stained cells in the acquired fluorescence images. Since the voxel sizes of acquired datasets were not isotropic, cell detection was performed on 2D images in the x-y plane with higher resolution and then stitched along the z direction. To improve model performance, a pretrained RetinaNet model on neurons in mouse brain datasets [29] was used as the starting point in the training process. Unique data augmentation methods, including brightness augmentation [28] and synthetic defocus according to physical modeling of point spread function [29,31], were also used. TeraStitcher [34] was employed to stitch multi-tile datasets, of which the output parameters were further used to stitch detected cells. Reconstructed data volumes were imported into Imaris 9.5 (Oxford Instruments) to trace the OC and segment the SG. Cochlear frequency maps were computed based on the traced OC according to Greenwood equation [14,35,36]. Corresponding frequencies of detected SGNs were estimated based on their shortest distances to the OC.

3.2 Data augmentation and transfer learning significantly improved cell detection performance

RetinaNet [30], a state-of-the-art object detection model, was trained to locate SGNs in the acquired fluorescence images. Given an input image, an object detection model is designed to output a list of targets of interest with corresponding class labels and coordinates of bounding boxes indicating the target locations. We have used RetinaNet to detect neurons and astrocytes in mouse brain fluorescence images for region-wise analysis and shown its superior performance [28,29]. The datasets used in this research included fluorescence images acquired from different samples with different markers imaged by two distinct LSFMs: gerbil datasets of TuJ1 staining, pig datasets of TuJ1 staining, PGP9.5 staining and autofluorescence. Images with ground truth annotations (bounding boxes around SGNs) were split into training data, validation data, and test data. To evaluate the model performance, average precision measure [37], a widely used evaluation metric for object detection networks, was utilized, which compares model predictions with ground truth annotations. Note that the average precision measure resembles the area under the curve (AUC) measure in a receiver operating characteristic (ROC) curve. We deployed unique data augmentation methods during training to improve the model performance, including brightness augmentation and synthetic defocus (Figure 2A).

Figure 2. Data augmentation and transfer learning improved model performance. (A) Spiral ganglion neuron (SGN) images with data augmentation strategies including saturation and/or synthetic defocus (scale bar, 25 µm). **(B)** Schematic illustration of transfer learning adopted in this research. A RetinaNet model was first trained on a real-world image dataset (COCO dataset [38]) and/or fluorescence images of mouse brain (MADM dataset [28,29]) for object detection. The acquired pretrained model was then used as the initial model for training on SGN dataset. **(C)** Comparison of model performance when training with transfer learning (COCO or MADM) and data augmentation (DA). Please note that the pretrained MADM model was also trained on the pretrained COCO model. Combining transfer learning and data augmentation significantly improved the model performance. *, $p < 0.05$; **, $p < 0.005$; unpaired t-tests, $n = 3$. **(D)** Precision-recall curve on unseen test data, showing an average precision of 0.872. **(E)** Representative images from different samples (gerbil and pig) and fluorescence labeling (TuJ1, PGP9.5, autofluorescence), with model predictions visualized as green bounding boxes (scale bar, 25 µm). Auto – autofluorescence.



A
Original
Saturated
Defocused
Defocused & saturated
B
COCO dataset
SGN dataset
MADM dataset
Pretrained model
SGN model
C
Model performance
Average precision (AP)
0.90
0.88
0.86
0.84
0.82
*
**
COCO
MADM
MADM+DA
D
Unseen test data
AUC = 0.872
Precision
Recall
1.0
0.5
0.0
0.0
0.5
1.0
E
Gerbil – TuJ1
Pig – TuJ1
Pig – PGP9.5
Pig – Auto

Data augmentation is widely used in deep learning to artificially increase the training data by applying transformations to existing data, e.g., flipping and rotation, which can expand the existing training data size without manual labeling of new images and therefore improve the model robustness. It was shown in our previous work [28] that data augmentation such as artificial saturation improved the model performance in cell detection. Therefore, to address variability in intensity level across different imaging depth within the samples, e.g., bright signal on the surface, while dimer signal deep in the tissue, we extended brightness augmentation to either saturate or desaturate the images. In addition, when the focal plane of objective lens is not perfectly overlapped with the illumination light sheet, images acquired from LSFM can seem out-of-focus. Out-of-focus images also happened when imaging with a standard Gaussian beam that is tight in the center of the field of view and diverge near the edges, thus causing a blur. Hence, we also synthetically defocused the in-focus training data based on physical modeling of microscope point spread functions [29,31] and included these synthetic images in the training data.

Transfer learning is commonly used when training models for complex tasks such as object detection, which can boost the model performance with small training dataset by employing knowledge of models pretrained on large datasets for related tasks [39]. We previously started with a pretrained model on real-world COCO dataset [38] and obtained a model showing excellent performance for cell detection in cleared MADM mouse brains [29]. Though the MADM model was trained to detect different types of cells (neurons and astrocytes) with different fluorescent labeling, its features were optimized on fluorescence images. Hence, we started our training for SGN detection with the pretrained MADM model (Figure 2B) and obtained a better average precision of 0.851 than 0.838 when starting with the pretrained COCO model (Figure 2C). Combining transfer learning with MADM model and unique data augmentation including brightness augmentation and synthetic defocus, model performance was significantly improved, achieving an average precision of 0.881 on validation data (Figure 2C).

The trained model with the best performance on validation data was selected as the inference model and evaluated on unseen test data (Table 1). An average precision of 0.872 was achieved on test data (Figure 2D). Specifically, the model achieved an average precision of 0.865 on 1632 TuJ1$^+$ cells in the gerbil datasets with a precision of 0.810 and a recall of 0.909 considering detections above confidence score of 0.5. On 1953 TuJ1$^+$ cells in the pig datasets, an average precision of 0.890 was achieved with a precision of 0.863 and a recall of 0.862. Trained on images with different fluorescent labeling, the model was able to detect cells with different morphologies from both gerbil and porcine cochlea samples (Figure 2E).

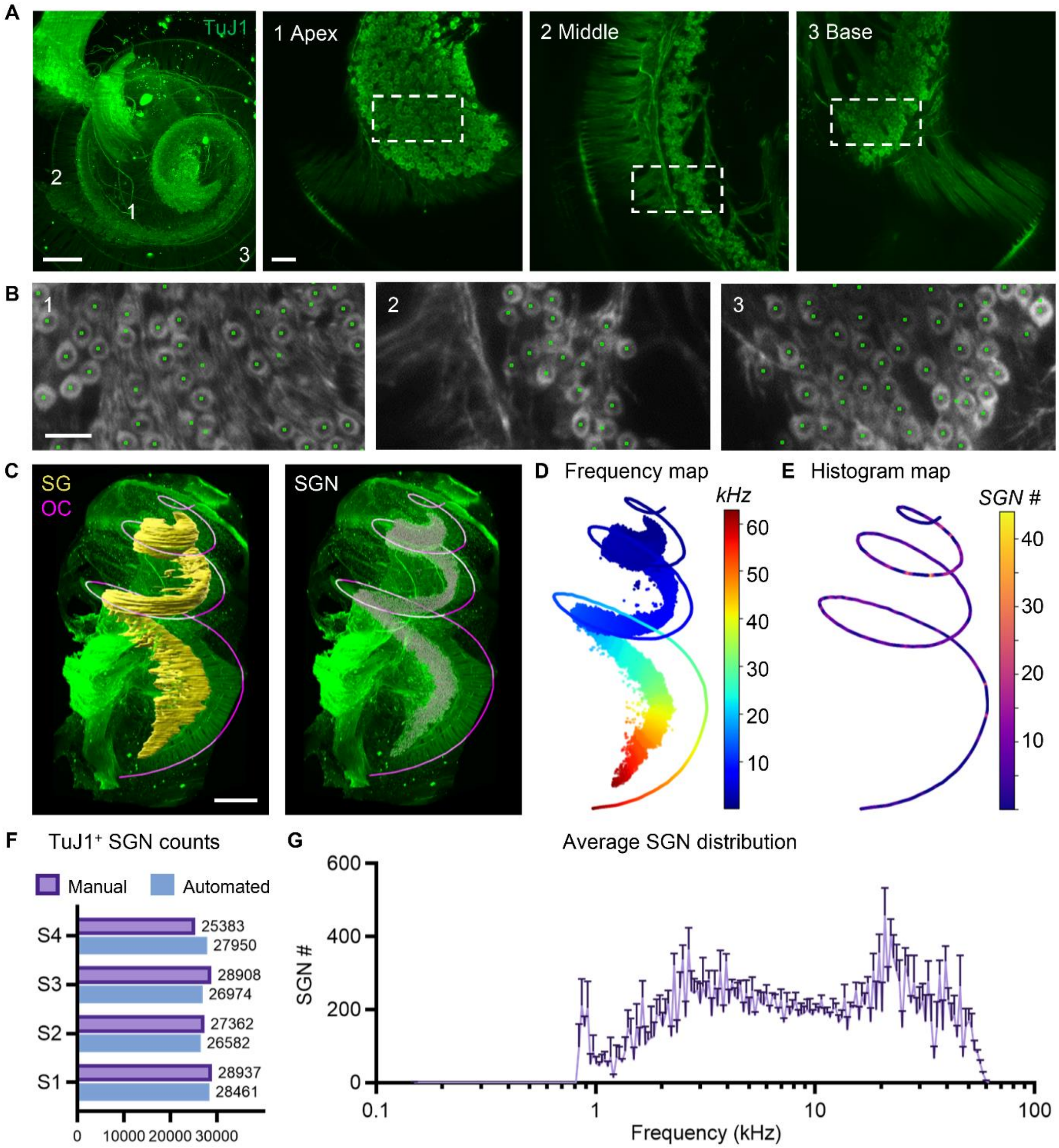
Table 1. Evaluation of model performance on unseen test data. Precisions, recalls, and F-scores were reported based on a confidence threshold of 0.5.

	Gerbil	Pig
# cells	1632	1953
Average precision	0.865	0.890
Precision (*0.5*)	0.810	0.863
Recall (*0.5*)	0.909	0.862
F-score *(0.5)*	0.857	0.862

3.3 Type I SGNs were accurately located by Spiner

To validate Spiner and the trained cell detection model, we applied them to gerbil cochleae with TuJ1 staining. TuJ1 has been reported to strongly express in Type I SGNs [40–43]. After tissue clearing, gerbil cochleae were imaged by a commercial LSFM, resulting in datasets with intact 3D structure and clear cellular information (Figure 3A and 3B). Applying the trained RetinaNet model to TuJ1 fluorescence images, individual SGNs were accurately distinguished at different locations near apex, middle, and base of the cochlea (Figure 3B). Please note that in Figure 3B, we have only indicated the centroids of the predicted bounding boxes. Placing the bounding boxes directly on the densely packed neurons would result in image overcrowding. Figure 3C shows the traced OC and segmented SG of the gerbil cochlea dataset, as well as detected SGNs. The cell detection model also detected small numbers of spiral ganglion cells seen outside the modiolus, adjacent to the nerve bundle (Figure S2). Given our focus on the SGN, all cells outside the SG region were excluded from the presented quantitative results. Cochlear frequency map along the OC was computed based on the Greenwood equation [35,36] (Figure 3D). Then corresponding frequencies of detected SGNs were estimated according to their shortest distances to the OC (Figure 3D). Afterwards, distribution of SGNs corresponding to the OC could be established (Figure 3E). Comparing automated cell counts of four samples with manual analysis, no significant differences were observed (Figure 3F), with a relative difference ranging from 1.64% to 10.11%, which demonstrated the efficacy and accuracy of the proposed workflow. Based on the estimated cochlear frequencies of detected SGNs, the histogram of SGN over the frequency map was obtained (Figure 3G). Note that it remains under debate the relation between apical SGNs and the OC.

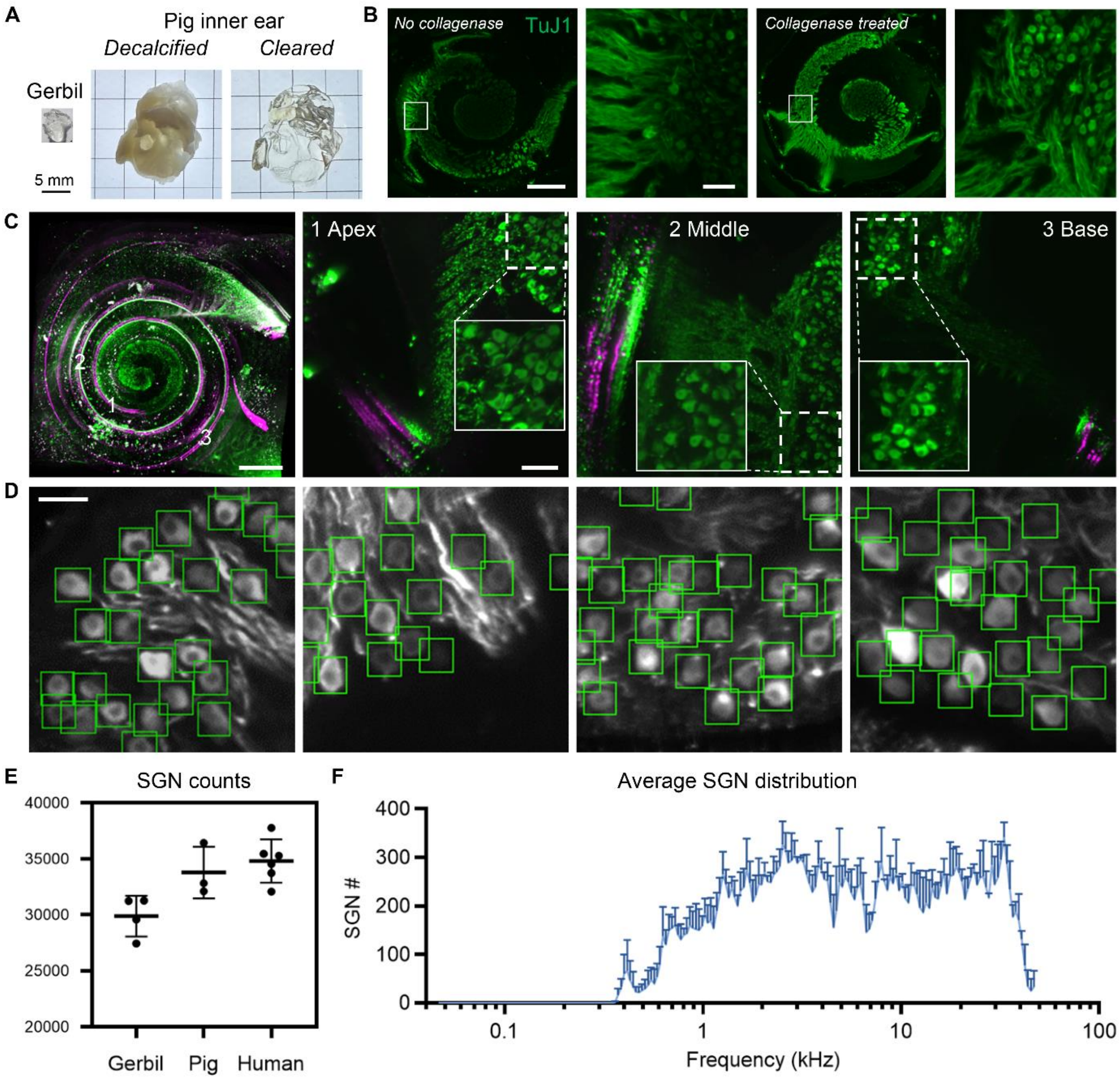
Figure 3. Type I SGNs were accurately located by Spiner. (A) Maximum intensity projection (scale bar, 200 μm) of TuJ1 fluorescence signals from a gerbil cochlea, with three zoomed in areas near apex, middle, and base of the cochlea (scale bar, 50 μm). **(B)** Network detections of spiral ganglion neurons corresponding to boxed areas in (A). Centroids of predicted bounding boxes were visualized as green dots. Scale bar, 25 μm. **(C)** Traced spiral ganglion (SG, yellow) and organ of Corti (OC, magenta) overlayed on the maximum intensity projection, as well as detected Type I spiral ganglion neurons (SGN, white dots). Detections out of the segmented spiral ganglion were excluded. Scale bar, 25 μm. **(D)** Cochlea frequency map along the OC was calculated based on the Greenwood equation [35,36]. Cochlear frequencies of the detected Type I SGNs were estimated given their shortest distances to the OC. **(E)** Corresponding counts of Type I SGNs along the OC. **(F)** Cell counts obtained from the automated workflow were close to manual analysis. **(G)** Average counts of detected Type I SGNs along the cochlear frequency (number of animals = 4, mean + SEM).


A
TuJ1
1 Apex
2 Middle
3 Base
2
1
3
B
1
2
3
C
SG
OC
SGN
D
Frequency map
kHz
60
50
40
30
20
10
E
Histogram map
SGN #
40
30
20
10
F
TuJ1+ SGN counts
Manual
Automated
S4
25383
27950
S3
28908
26974
S2
27362
26582
S1
28937
28461
0
10000
20000
30000
G
Average SGN distribution
SGN #
600
400
200
0
0.1
1
10
100
Frequency (kHz)

3.4 Spiner successfully mapped Type I SGNs in the porcine cochlea

We then applied Spiner to tissue clear, image, and analyze porcine cochleae. Our previous work showed successful staining of hair cells in whole-mount porcine cochlea samples [13,14,16]. Compared with gerbil cochlea, the size of porcine cochlea is much larger (Figure 4A and Figure S3A), making it more challenging for whole-mount antibody staining of SGNs that are embedded inside the bone in the cochlea. We optimized the BoneClear [4] protocol and added a collagenase treatment procedure before permeabilization to improve the antibody penetration. As shown in Figure 4B, a more uniform antibody staining was observed with collagenase treatment while the structural information and cellular morphology remained the same (see also Figure S3B). We validated TuJ1 staining and achieved co-staining of hair cells and SGNs with consistent good signals from apex to the base (Figure 4C). Imaging with a custom-built LSFM [14,32,33], we obtained decent image quality across the entire sample. Given the large size of the porcine cochlea, the entire cochlea could not be captured using a single tile, and therefore 5 by 5 tiles with 15 % overlap were acquired. The overlap was used to later stitch the different tiles to a 3D volume.

Figure 4. Spiner successfully mapped Type I SGNs in the porcine cochlea. (A) A 6–8-week-old pig's inner ear before and after tissue clearing, compared with a cleared gerbil cochlea (scale bar, 5 mm). The samples were imaged in the air. **(B)** Effect of collagenase treatment on TuJ1 antibody penetration. Virtual slices of 10 µm thickness from samples without or with collagenase treatment are shown (scale bar, 500 µm), with zoomed in images of the boxed areas (scale bar, 50 µm). **(C)** Maximum intensity projection (scale bar, 1 mm) of fluorescence signals from a porcine cochlea showing co-staining of Myo7a (magenta) for hair cells and TuJ1 (green) for Type I SGNs (scale bar, 1 mm), with three zoomed in areas near apex, middle, and base of the cochlea (scale bar, 100 µm). **(D)** Representative images of SGN predictions (green bounding boxes) from multiple pig cochlea datasets (scale bar, 25 µm). **(E)** Comparison of SGN counts (mean ± SD) among gerbil, pig, and human [44]. For gerbil and pig, the whole SGN counts were estimated assuming 92.5% SGNs are Type I SGNs [45]. **(F)** Average counts of detected Type I SGNs along the cochlear frequency (number of animals = 3, mean + SEM).



A
Pig inner ear
Decalcified
Cleared
Gerbil
5 mm
B
No collagenase
TuJ1
Collagenase treated
C
1 Apex
2 Middle
3 Base
D
E
SGN counts
Gerbil
Pig
Human
F
Average SGN distribution
SGN #
Frequency (kHz)

The trained RetinaNet model was able to detect stained SGNs in the porcine cochlea datasets accurately (Figure 4D). After detecting the cells in each image tile, the cells were stitched in 3D using TeraStitcher [34] parameters; cells that appear in two adjacent and overlapping tiles were excluded [29]. We analyzed three porcine samples and obtained an average Type I SGN count of 29627 ± 1696. Given 90% to 95% of SGNs [45] are Type I SGNs, we estimated the total numbers of SGNs in the porcine cochlea as 32030 ± 1834, which is comparable to human SGN counts of 34817 ± 1911 [44] (Figure 4E). Similarly, we estimated the cochlea frequencies of detected SGNs and obtained the histogram of cell counts over the frequency map (Figure 4F). In summary, using the proposed workflow, we were able to tissue clear porcine cochlea with excellent quality of TuJ1 staining and accurately locate stained Type I SGNs in the porcine cochlea with consistent cell counts across multiple samples. Note, similarly to the gerbil, stained TuJ1 cells were observed adjacent to the nerve bundle, these neurons were excluded from our counts.

4 Discussion

In this manuscript, we present a workflow to tissue clear, label, image and analyze entire cochlea in 3D. We named the workflow – Spiner. Using Spiner, we successfully stained and counted the Type I SGNs in the gerbil and porcine cochlea. By adding a collagenase treatment procedure into the tissue clearing protocol, we achieved uniform staining of TuJ1 in the large pig cochlea. After, we trained a RetinaNet model for SGN detection and achieved an average precision of 0.872 by utilizing transfer learning approach and data augmentation with physical modeling of defocus introduced. The RetinaNet model was robust to detect SGNs with different markers (TuJ1 and PGP9.5), even in autofluorescence. We validated the proposed workflow on gerbil datasets and obtained consistent cell counts close to manual analysis, with a relative difference ranging from 1.64% to 10.11%. Type I SGNs in porcine cochlea were located and quantified, counts of which were comparable to those of human SGNs, further showing the efficacy of pig as a large animal model in the auditory research. The proposed workflow can be utilized in future research to study the effects of cochlear implants or therapeutic treatments on SGNs by quantification of the SGN number and survival in 3D.

In the future, the method will be extended to segmentation of processes from the SGNs to the OC, which will facilitate accurate analysis of the functional frequencies of SGNs and provide valuable insights of studying SGN innervation to the hair cells, thus offering a comprehensive approach to quantitative analysis of SGNs in entire cochlea samples.

We introduced unique data augmentation of brightness augmentation and synthetic defocus, which significantly improved the model performance on cell detection. In the future, other types of physical aberration models can be included in the data augmentation, which can further improve the robustness of deep learning models on various image configurations.

The generalization of deep learning models is always challenging, especially when applying it to different samples or different imaging instruments. Retraining the model usually requires extensive effort and labor to annotate training data. Alternatively, generative networks such as generative adversarial networks [46] and diffusion networks [47] can be used to generate synthetic training data with known ground truth and minimal human labor. Compared to generative adversarial networks, diffusion networks have the advantages of adding constraints during data generation. We have tested the capability of diffusion models to generate fluorescence images of SGNs that are undistinguishable from real images (Figure S4). Therefore, it is worth exploring the use of diffusion networks when generalizing trained models to new datasets.

3D microscopy datasets are not always imaged with isotropic resolution. In most cases, the resolution in z-axis is lower, which makes cell detection in 3D troublesome. Additionally, it is more difficult to annotate objects in 3D than in 2D. In this research, we first detected cells in high-resolution 2D images, yet information in the z-axis was roughly used to estimate the z location of each object. In the future, research can be done to locate the objects in 3D more accurately, either using algorithms to determine the boundaries of objects across z-planes or integrating regression-type density maps to locate cell locations in the z-axis.

In this manuscript, we estimated the cochlear frequencies of detected SGNs by measuring their shortest distances to the traced OC, which is not very accurate and couldn’t be used to resolve the frequencies near the apex of the cochlea. Alternatively, frequency assignment for SGNs can be performed by matching the percentage location of SGNs along the SG to the OC [48]. The optimal solution is to trace the nerve fibers, yet it is challenging in 3D datasets given the fiber densities. Note that it is also under debate that if the Greenwood [36] equation is an appropriate mapping of cochlear frequency, especially when near the apex of the cochlea [49].

5. Materials and methods

5.1 Sample details

Porcine cochleae were extracted from wild-type Yorkshire pigs aging from embryonic day 80 to 8 weeks.

5.2 Tissue clearing porcine tissue

BoneClear [4] protocol was modified to stain and clear porcine cochleae. The extracted cochlea samples were post-fixed in 0.5% Paraformaldehyde (PFA) and 10% sucrose in PBS for 2 h at room temperature (RT), and further fixed in 0.5% PFA in PBS overnight at 4 °C. The fixed samples were washed three times in PBS, 1 h each at RT. The cochleae were decalcified in 350 mM EDTA-Na (pH 6.5) for 3-7 days at 37 °C, depending on the sample age and size, with a fresh solution changed every day. The decalcified samples were dehydrated using methanol gradients (diluted in ddH_2O): 20%, 40%, 60%, 80%, 100%, 100%; 2 h each at RT. The samples were then decolorized overnight at 4 °C in a chilled fresh mixture of 30% H_2O_2 and methanol (v:v=1:10). The bleached samples were rehydrated with the inverse methanol gradients (diluted in ddH_2O): 100%, 80%, 40%, 20%; 2 h each at RT. The samples were then washed with PBS for 2 h at RT. The samples were then treated with 0.2% (w/v) Collagenase in PBS for 45 min at 37°C and washed twice for 10 min with 2% fetal bovine serum in PBS at 37°C. The samples were then permeabilized with PBS/0.2% Triton X-100/0.1% Deoxycholate/10% DMSO/25 mM EDTA (pH 6.5) for 2 d at 37 °C and then blocked with PBS/0.2% Triton X-100 /10% DMSO/5% donkey serum/25 mM EDTA (pH 6.5) overnight at 37 °C. The tissues were then incubated with rabbit anti-Myo7a (1:200, Proteus Biosciences) and/or mouse anti-TuJ1 (1:100, BioLegend) diluted in the antibody buffer made of PBS/0.2% Tween-20/10 µg/mL heparin/5% normal donkey serum/25 mM EDTA (pH 6.5) for 7 d at 37°C. After primary staining, the samples were washed with the washing buffer made of PBS/0.2% Tween-20/10 µg/mL heparin/25 mM EDTA (pH 6.5) for 24 h at 37°C, with a fresh buffer changed every 6 h. The samples were then stained with donkey anti-rabbit Cy3 (1:200, Jackson ImmunoResearch) and/or donkey anti-mouse AlexaFluor 647 (1:200, Jackson ImmunoResearch) in the antibody buffer for 5 d at 37°C. The samples were then washed with the washing buffer for 48 h at 37°C, with a fresh buffer changed every 6 h. For optical clearing, the cochlea samples were dehydrated in 20% for 4 h at RT and then dehydrated with methanol gradients diluted in ddH_2O: 40%, 60%, 80%, 100%, 100%; 2 h each at RT. The samples were then incubated in the mixture of dichloromethane and methanol (v:v=2:1) for 2 h twice at RT, followed by incubation in dichloromethane for 30 min four times at RT. The cochleae were then cleared with dibenzyl ether for 12 h three times at RT. All the incubation steps were performed with gentle shaking.

5.3 Light-sheet imaging

Cleared porcine cochleae were imaged in dibenzyl ether (refractive index 1.564) using a custom-built light sheet fluorescence microscope as summarized in [14,32,33], by illumination of 561 nm and 640 nm respectively, resulting in datasets with a voxel size of $0.65 \times 0.65 \times 10\ \mu m^3$. The acquired multi-tile datasets were stitched using TeraStitcher [34].

5.4 Data preparation

The annotated data contained fluorescence images from both gerbil cochlea datasets stained for TuJ1 and porcine cochlea datasets including images of autofluorescence, PGP9.5 staining and TuJ1 staining. 10045 individual cells were labeled in the training data and 3644 individual cells were labeled in the validation data. Unseen test data were generated from samples different than those that were used to prepare training data and validation data. 4448 cells (1632 from gerbil datasets and 2816 from porcine datasets) were labeled in the test data.

5.5 Cell detection

A RetinaNet [30] model adapted from the source repository (https://github.com/fizyr/keras-retinanet) was trained to locate spiral ganglion neurons in the images. Transfer learning was adopted with a pre-trained RetinaNet model [29] on fluorescence microscopy images being used as the initial model for training. Data augmentation strategies such as geometrical transforms and brightness augmentation were applied [28], as well as including synthetic defocused training data generated based on physical modeling [31]. The initial learning rate was 0.0001 and the batch size was eight. The model was trained for 50 epochs with an Adam optimizer using Nvidia Tesla V100 GPU on the UNC longleaf cluster. Average precision measure [37] was used to comprehensively evaluate model performance, which is widely used in object detection tasks; it could be referred to as the area under the precision-recall curve. The model showing the best performance on validation data was evaluated on test data and used as the inference model. Precisions, recalls, and F-scores [50] were reported based on a confidence threshold of 0.5.

For multi-tile datasets, cell detection was first performed per tile of 2D images, during which images were cropped into image patches of 512×512 with an overlap fraction of 0.125 for inference. The detected cells in each tile were then stitched in 3D according to TeraStitcher [34] parameters. Non-maximum suppression was performed to remove duplicate detections [29].

Given different imaging step sizes, a cell could be detected in multiple z-planes (Figure S2A). Therefore, along the z-axis, the overlapped detections, determined by an intersection over union above 0.5, were merged based on the observed cell occurrence across z-planes in each dataset.

Depending on the data size, the cell detection process takes around three hours per gerbil dataset and around eight hours per porcine dataset.

5.6 Tracing of organ of Corti and spiral ganglion

The 3D data volume and detected cells was imported into Imaris 9.5 (Oxford Instruments) for visualization and post-processing.

The “Measurement Points” tool was used to manually add points about every 60 µm along the organ of Corti across z-planes. After tracing, the 3D coordinates of assigned points were exported to compute cochlear frequency maps.

The “Surfaces” tool was used to manually segment the spiral ganglion by drawing contours of region of interests onto z-planes. The acquired 3D segmentation mask was then exported into MATLAB for further analysis.

5.7 Cochlear frequency estimation

Cochlear frequency maps were computed based on the Greenwood equation [14,35,36]. For gerbil cochlea, $CF(pos) = 0.398 \cdot (10^{(100-pos)\cdot 0.022} - 0.631)$, where CF is the cochlear frequency in kHz and pos is the percentage basilar membrane length from the base of the cochlea [35]. For porcine cochlea, $CF(pos) = 0.1654 \cdot 2.3 \cdot (10^{(100-pos)\cdot 0.021} - 0.88)$ [14,36]. Cochlear frequencies of spiral ganglion neurons were estimated based on their shortest distances to the cochlear frequency map along the traced organ of Corti.

5.8 Statistical analysis

Statistical analysis was performed using Prism 10 (GraphPad Software).

REFERENCES

1. Deafness and hearing loss [Internet]. [cited 2024 Jan 29]. Available from: https://www.who.int/health-topics/hearing-loss
2. Lv J, Wang H, Cheng X, Chen Y, Wang D, Zhang L, et al. AAV1-hOTOF gene therapy for autosomal recessive deafness 9: a single-arm trial. The Lancet [Internet]. 2024 Jan 24 [cited 2024 Feb 2];0(0). Available from: https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(23)02874-X/fulltext
3. Davis RL. Gradients of Neurotrophins, Ion Channels, and Tuning in the Cochlea. Neuroscientist. 2003 Oct 1;9(5):311–6.
4. Wang Q, Liu K, Yang L, Wang H, Yang J. BoneClear: whole-tissue immunolabeling of the intact mouse bones for 3D imaging of neural anatomy and pathology. Cell Res. 2019 Oct;29(10):870–2.
5. Ueda HR, Ertürk A, Chung K, Gradinaru V, Chédotal A, Tomancak P, et al. Tissue clearing and its applications in neuroscience. Nat Rev Neurosci. 2020 Feb;21(2):61–79.
6. Richardson DS, Lichtman JW. Clarifying Tissue Clearing. Cell. 2015 Jul;162(2):246–57.
7. Renier N, Adams EL, Kirst C, Wu Z, Azevedo R, Kohl J, et al. Mapping of Brain Activity by Automated Volume Analysis of Immediate Early Genes. Cell. 2016 Jun 16;165(7):1789–802.
8. Zhao S, Todorov MI, Cai R, -Maskari RA, Steinke H, Kemter E, et al. Cellular and Molecular Probing of Intact Human Organs. Cell. 2020 Feb 20;180(4):796-812.e19.
9. Biswas L, Chen J, Angelis JD, Chatzis A, Nanchahal J, Dustin ML, et al. SUMIC: A Simple Ultrafast Multicolor Immunolabelling and Clearing Approach for Whole-Organ and Large Tissue 3D Imaging [Internet]. bioRxiv; 2021 [cited 2024 Jan 29]. p.

2021.01.20.427385. Available from:
https://www.biorxiv.org/content/10.1101/2021.01.20.427385v1

10. Nudell V, Wang Y, Pang Z, Lal NK, Huang M, Shaabani N, et al. HYBRiD: hydrogel-reinforced DISCO for clearing mammalian bodies. Nat Methods. 2022 Apr;19(4):479–85.

11. Cai R, Pan C, Ghasemigharagoz A, Todorov MI, Förstera B, Zhao S, et al. Panoptic imaging of transparent mice reveals whole-body neuronal projections and skull–meninges connections. Nat Neurosci. 2019 Feb;22(2):317–27.

12. Voie AH, Burns DH, Spelman FA. Orthogonal-plane fluorescence optical sectioning: Three-dimensional imaging of macroscopic biological specimens. Journal of Microscopy. 1993;170(3):229–36.

13. Moatti A, Li C, Sivadanam S, Cai Y, Ranta J, Piedrahita JA, et al. Ontogeny of cellular organization and LGR5 expression in porcine cochlea revealed using tissue clearing and 3D imaging. iScience. 2022 Aug 19;25(8):104695.

14. Moatti A, Cai Y, Li C, Sattler T, Edwards L, Piedrahita J, et al. Three-dimensional imaging of intact porcine cochlea using tissue clearing and custom-built light-sheet microscopy. Biomed Opt Express, BOE. 2020 Nov 1;11(11):6181–96.

15. Hutson KA, Pulver SH, Ariel P, Naso C, Fitzpatrick DC. Light Sheet Microscopy of the Gerbil Cochlea. J Comp Neurol. 2021 Mar;529(4):757–85.

16. Moatti A, Cai Y, Li C, Popowski KD, Cheng K, Ligler FS, et al. Tissue clearing and three-dimensional imaging of the whole cochlea and vestibular system from multiple large-animal models. STAR Protocols. 2023 Jun 16;4(2):102220.

17. MacDonald GH, Rubel EW. Three-dimensional imaging of the intact mouse cochlea by fluorescent laser scanning confocal microscopy. Hear Res. 2008 Sep;243(1–2):1–10.

18. Urata S, Iida T, Yamamoto M, Mizushima Y, Fujimoto C, Matsumoto Y, et al. Cellular cartography of the organ of Corti based on optical tissue clearing and machine learning. King AJ, Whitfield TT, Heller S, editors. eLife. 2019 Jan 18;8:e40946.

19. Johnson SB, Schmitz HM, Santi PA. TSLIM imaging and a morphometric analysis of the mouse spiral ganglion. Hearing Research. 2011 Aug 1;278(1):34–42.

20. Wrzeszcz A, Reuter G, Nolte I, Lenarz T, Scheper V. Spiral ganglion neuron quantification in the guinea pig cochlea using Confocal Laser Scanning Microscopy compared to embedding methods. Hearing Research. 2013 Dec;306:145–55.

21. Buswinka CJ, Osgood RT, Simikyan RG, Rosenberg DB, Indzhykulian AA. The hair cell analysis toolbox is a precise and fully automated pipeline for whole cochlea hair cell quantification. PLOS Biology. 2023 Mar 22;21(3):e3002041.

22. Cortada M, Sauteur L, Lanz M, Levano S, Bodmer D. A deep learning approach to quantify auditory hair cells. Hearing Research. 2021 Sep 15;409:108317.

23. Echteler SM, Nofsinger Y c. Development of ganglion cell topography in the postnatal cochlea. Journal of Comparative Neurology. 2000;425(3):436–46.

24. Richter CP, Kumar G, Webster E, Banas SK, Whitlon DS. Unbiased counting of neurons in the cochlea of developing gerbils. Hearing Research. 2011 Aug 1;278(1):43–51.

25. Pan C, Schoppe O, Parra-Damas A, Cai R, Todorov MI, Gondi G, et al. Deep Learning Reveals Cancer Metastasis and Therapeutic Antibody Targeting in the Entire Body. Cell. 2019 Dec;179(7):1661-1676.e19.

26. Todorov MI, Paetzold JC, Schoppe O, Tetteh G, Shit S, Efremov V, et al. Machine learning analysis of whole mouse brain vasculature. Nat Methods. 2020 Apr;17(4):442–9.

27. Sità L, Brondi M, Lagomarsino de Leon Roig P, Curreli S, Panniello M, Vecchia D, et al. A deep-learning approach for online cell identification and trace extraction in functional two-photon calcium imaging. Nat Commun. 2022 Mar 22;13(1):1529.

28. Cai Y, Zhang X, Kovalsky SZ, Ghashghaei HT, Greenbaum A. Detection and classification of neurons and glial cells in the MADM mouse brain using RetinaNet. PLOS ONE. 2021 Sep 24;16(9):e0257426.

29. Cai Y, Zhang X, Li C, Ghashghaei HT, Greenbaum A. COMBINe enables automated detection and classification of neurons and astrocytes in tissue-cleared mouse brains. Cell Reports Methods. 2023 Apr 24;3(4):100454.

30. Lin TY, Goyal P, Girshick R, He K, Dollár P. Focal Loss for Dense Object Detection [Internet]. arXiv; 2018 [cited 2023 Mar 15]. Available from: http://arxiv.org/abs/1708.02002

31. Yang SJ, Berndl M, Michael Ando D, Barch M, Narayanaswamy A, Christiansen E, et al. Assessing microscope image focus quality with deep learning. BMC Bioinformatics. 2018 Mar 15;19(1):77.

32. Li C, Moatti A, Zhang X, Ghashghaei HT, Greenbaum A. Deep learning-based autofocus method enhances image quality in light-sheet fluorescence microscopy. Biomed Opt Express, BOE. 2021 Aug 1;12(8):5214–26.

33. Li C, Rai MR, Ghashghaei HT, Greenbaum A. Illumination angle correction during image acquisition in light-sheet fluorescence microscopy using deep learning. Biomed Opt Express, BOE. 2022 Feb 1;13(2):888–901.

34. Bria A, Iannello G. TeraStitcher - A tool for fast automatic 3D-stitching of teravoxel-sized microscopy images. BMC Bioinformatics. 2012 Nov 27;13(1):316.

35. Müller M. The cochlear place-frequency map of the adult and developing mongolian gerbil. Hearing Research. 1996 May 1;94(1):148–56.

36. Greenwood DD. A cochlear frequency-position function for several species—29 years later. The Journal of the Acoustical Society of America. 1990 Jun 1;87(6):2592–605.

37. Everingham M, Eslami SMA, Van Gool L, Williams CKI, Winn J, Zisserman A. The Pascal Visual Object Classes Challenge: A Retrospective. Int J Comput Vis. 2015 Jan;111(1):98–136.

38. Lin TY, Maire M, Belongie S, Bourdev L, Girshick R, Hays J, et al. Microsoft COCO: Common Objects in Context. arXiv:14050312 [cs] [Internet]. 2015 Feb 20 [cited 2019 Dec 9]; Available from: http://arxiv.org/abs/1405.0312

39. Pan SJ, Yang Q. A Survey on Transfer Learning. IEEE Transactions on Knowledge and Data Engineering. 2010 Oct;22(10):1345–59.

40. Liu W, Kinnefors A, Boström M, Rask-Andersen H. Expression of peripherin in human cochlea. Cell Tissue Res. 2010 Dec 1;342(3):345–51.

41. Sekerková G, Zheng L, Mugnaini E, Bartles JR. Espin actin-cytoskeletal proteins are in rat type I spiral ganglion neurons and include splice-isoforms with a functional nuclear localization signal. J Comp Neurol. 2008 Aug 20;509(6):661–76.

42. Liu W, Boström M, Rask-Andersen H. Expression of peripherin in the pig spiral ganglion – aspects of nerve injury and regeneration. Acta Oto-Laryngologica. 2009 Jan 1;129(6):608–14.

43. Barclay M, Ryan AF, Housley GD. Type I vs type II spiral ganglion neurons exhibit differential survival and neuritogenesis during cochlear development. Neural Development. 2011 Oct 11;6(1):33.

44. Hinojosa R, Nelson EG. Cochlear nucleus neuron analysis in individuals with presbycusis. The Laryngoscope. 2011;121(12):2641–8.

45. Sanders TR, Kelley MW. Specification of neuronal subtypes in the spiral ganglion begins prior to birth in the mouse. Proceedings of the National Academy of Sciences. 2022 Nov 29;119(48):e2203935119.

46. Hollandi R, Szkalisity A, Toth T, Tasnadi E, Molnar C, Mathe B, et al. nucleAIzer: A Parameter-free Deep Learning Framework for Nucleus Segmentation Using Image Style Transfer. Cell Systems. 2020 May 20;10(5):453-458.e6.

47. Xu X, Guo J, Wang Z, Huang G, Essa I, Shi H. Prompt-Free Diffusion: Taking “Text” out of Text-to-Image Diffusion Models [Internet]. arXiv; 2023. Available from: https://arxiv.org/abs/2305.16223

48. Stakhovskaya O, Sridhar D, Bonham BH, Leake PA. Frequency Map for the Human Cochlear Spiral Ganglion: Implications for Cochlear Implants. J Assoc Res Otolaryngol. 2007 Jun;8(2):220–33.

49. Walia A, Ortmann AJ, Lefler S, Holden TA, Puram SV, Herzog JA, et al. Place Coding in the Human Cochlea [Internet]. medRxiv; 2023 [cited 2024 Jan 29]. p. 2023.04.13.23288518. Available from: https://www.medrxiv.org/content/10.1101/2023.04.13.23288518v2

50. Hossin M, Sulaiman MN. A Review on Evaluation Metrics for Data Classification Evaluations. International Journal of Data Mining & Knowledge Management Process. 2015 Mar 31;5:01–11.

CHAPTER 5: Conclusion and future work

In this dissertation, we implemented an object detection network named RetinaNet to detect MADM-labeled neurons and glia in fluorescence images acquired from mouse brain sections and achieved an average precision of 0.90. Additionally, we introduced a novel data augmentation method aimed at mitigating color-, intensity-, and saturation-related biases present in the dataset resulting from the investigated genotypes and the imaging conditions. To accurately detect and count individual glial cells within dense glia clusters that showed high variability in their shapes and fluorophore densities, a second RetinaNet model was dedicated. Integrating predictions from both models significantly enhanced automated cell counting within glia clusters. This cell detection method can be applied to various studies requiring quantification of labeled cells in fluorescence images.

We then developed COMBINe to locate and classify neurons and astrocytes in entire tissue cleared MADM mouse brains. The trained RetinaNet was robust to detect slightly out-of-focus cells in the acquired images. Utilizing COMBINe, processing and analysis of a dataset containing nearly 400 GB images and half a million labeled cells was completed in four days, which can take months for tedious manual quantification. By registering the reconstructed data volumes and detected cells to Allen Brain Atlas, we quantitatively analyzed the effects of MADM-based deletion of Egfr on neuron and astrocyte populations during the developmental stage and observed significant changes in astrocyte population consistent with laborious manual analysis, which validated the efficacy of the proposed approach. COMBINe is adaptable for diverse research utilizing tissue clearing and immunostaining of entire mouse brains.

In addition, a deep learning-based workflow named Spiner was developed to detect and quantify SGNs in entire gerbil and porcine cochleae. By adding a collagenase treatment to improve antibody penetration, uniform antibody staining of SGNs within the modiolus can be achieved in porcine cochlea whose size is much larger than rodent cochlea. The trained RetinaNet model was robust to SGN datasets acquired from different sample preparations labeled with different markers. We quantified Type I SGNs in the gerbil cochlea datasets and obtained cell counts that were close to manual counting which further validated the proposed method. Type I SGNs of porcine cochlea was also reported. Correlating the SGNs to cochlear frequency maps, distribution of SGNs could be compared across samples. This approach provides a robust tool for studying and quantifying SGNs in cochlea, advancing the study of auditory neural functions for potential clinical applications.

To summarize, these deep learning-based cell detection approaches hold promise for diverse fluorescence image datasets, empowering researchers with valuable tools for comprehensive image analysis and advancing the understanding of complex biological systems.

For future endeavors, expanding the cell detection process from 2D to 3D is a priority. Given the limited axial resolution along the z-axis, cell detection in this dissertation was confined to 2D images using RetinaNet. While achieving high accuracy, redundant cell detection across multiple z-planes posed challenges during post-processing. Transitioning to 3D cell detection would improve the accuracy of cell localization along the z-direction.

Moreover, integrating active learning could enhance the applicability of proposed methods to unseen datasets. Through iterative training of pretrained models with additional annotations from human experts, model performance could be optimized when adapting to new datasets while alleviating the annotation burden.

Addressing the need for generalizing models across diverse datasets is both time-consuming and resource-intensive due to the preparation of new training data with ground truth annotations. To tackle this challenge, generative networks like generative adversarial networks and diffusion networks offer promise in synthesizing training data with ground truth. These networks have demonstrated the ability to generate biomedical images indistinguishable from real data, thus offering potential for synthesizing training data to adapt pretrained models to new datasets while minimizing the human effort required for annotation.

APPENDICES

Appendix A: CHAPTER 2 Supplemental information

1 Supplemental figures

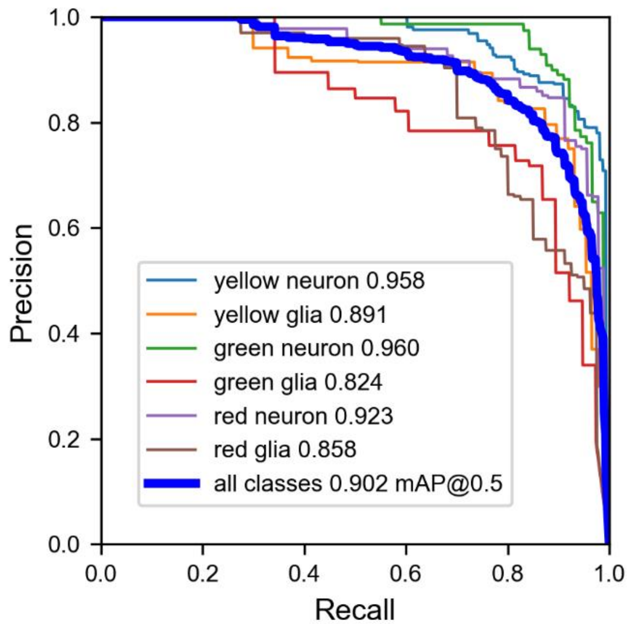
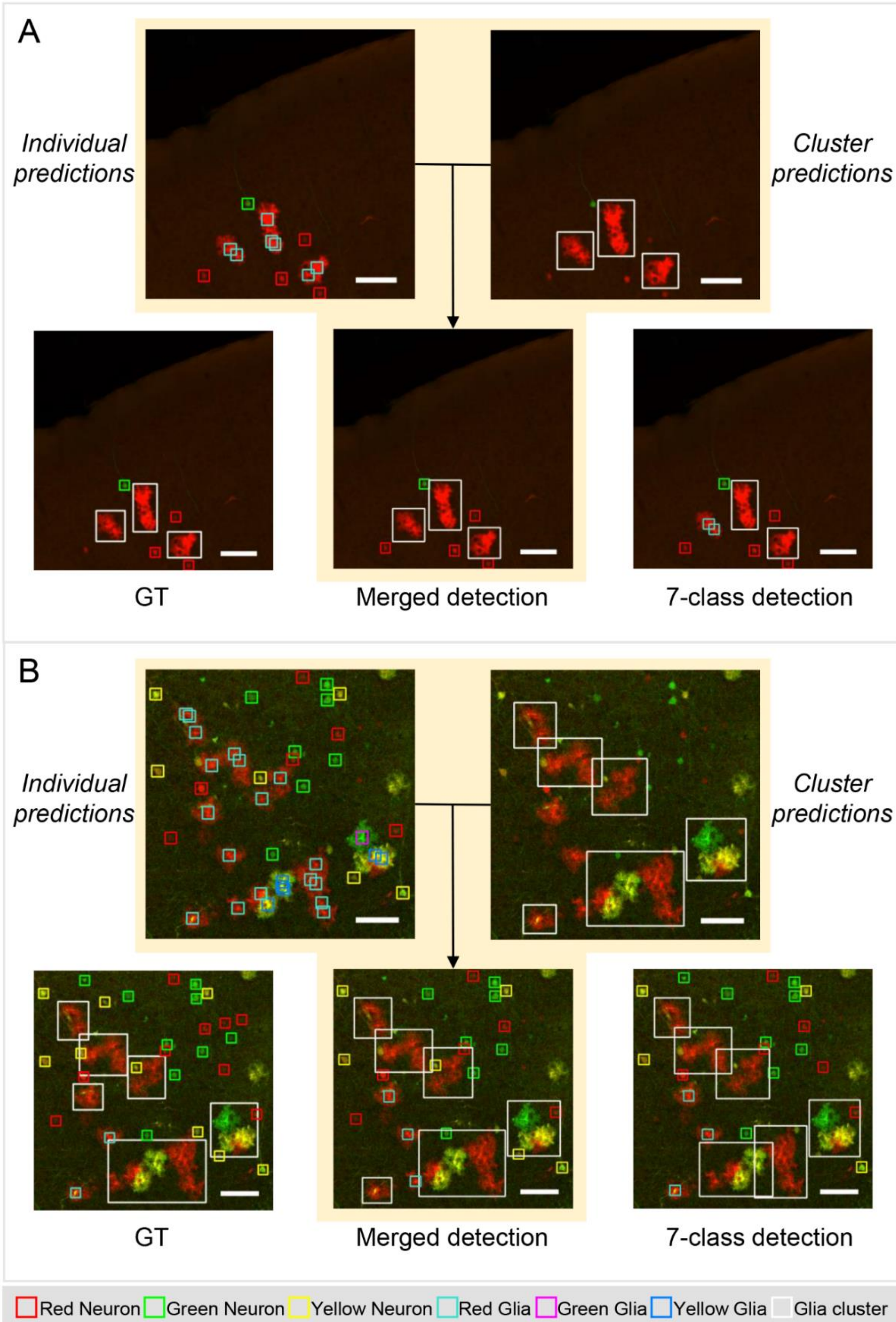


Figure S1. Precision-recall curves across six classes in individual cell detection.

Figure S2. Examples of merging results compared to the ground truth (GT) annotations. **(A, B)** Representative merging results of the images acquired from the slide scanner and the confocal fluorescence microscope (CFM), respectively. Two RetinaNet models were trained separately: one to detect individual cells and one to detect glia clusters (Figure 1E). Predictions of individual cells and glia clusters were then merged to evaluate the performance. For comparison, a RetinaNet model was trained to detect seven classes simultaneously (red/green/yellow neuron, red/green/yellow glia, and glia cluster). Predictions on the same image patches with confidence above 0.5 are shown. Note that based on the merging rules, cluster predictions with confidence above 0.3 are also considered in the merging process. Scale bars, 100 µm.


A
Individual predictions
Cluster predictions
GT
Merged detection
7-class detection
B
Individual predictions
Cluster predictions
GT
Merged detection
7-class detection
Red Neuron
Green Neuron
Yellow Neuron
Red Glia
Green Glia
Yellow Glia
Glia cluster

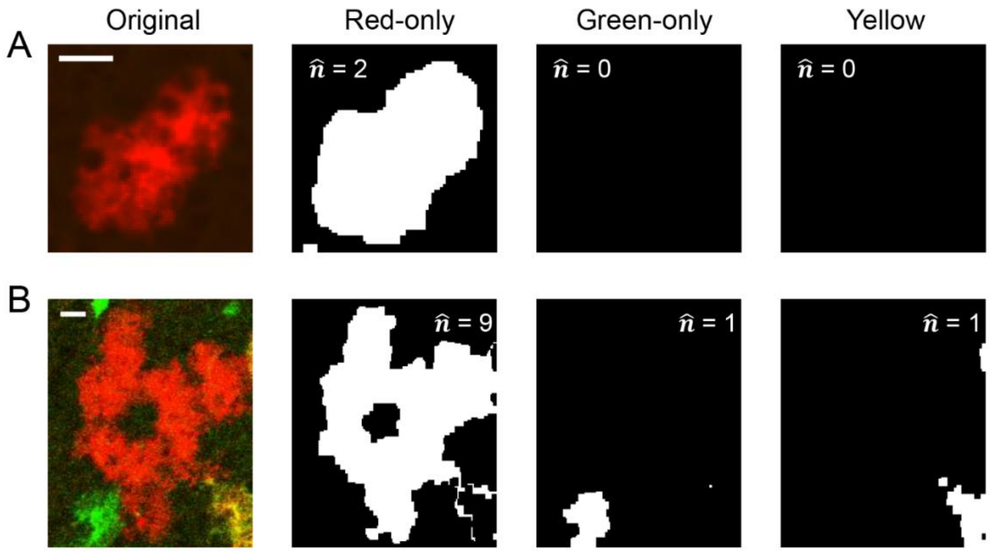


Figure S3. Representative area-based counting within glia clusters. (A, B) Counting results of glia clusters from images acquired using a slide scanner and a CFM, respectively. Binary masks of cells regions were generated for each color by thresholding and morphological operations. Estimated cell numbers of each color are marked in the images. The glia cluster in A contains 2 red glia and the glia cluster in B contains 9 red glia, 2 green glia and 1 yellow glial cell according to ground truth annotations. Scale bars, 25 µm.

Appendix B: CHAPTER 3 Supplemental information

1 Supplemental figures

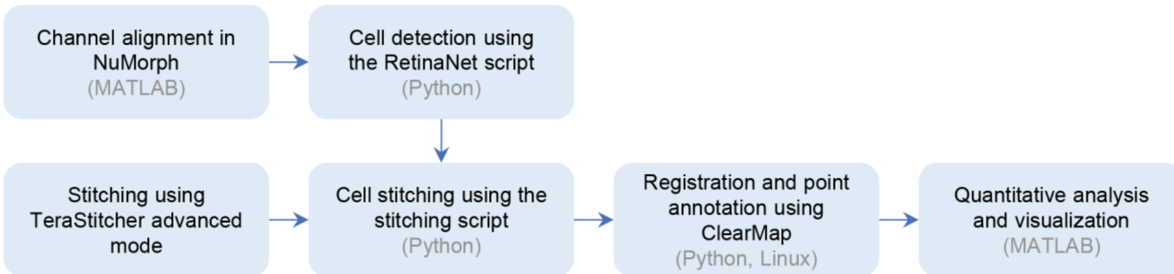


Figure S1. Running the pipeline for MADM cell mapping. Related to Figure 1.

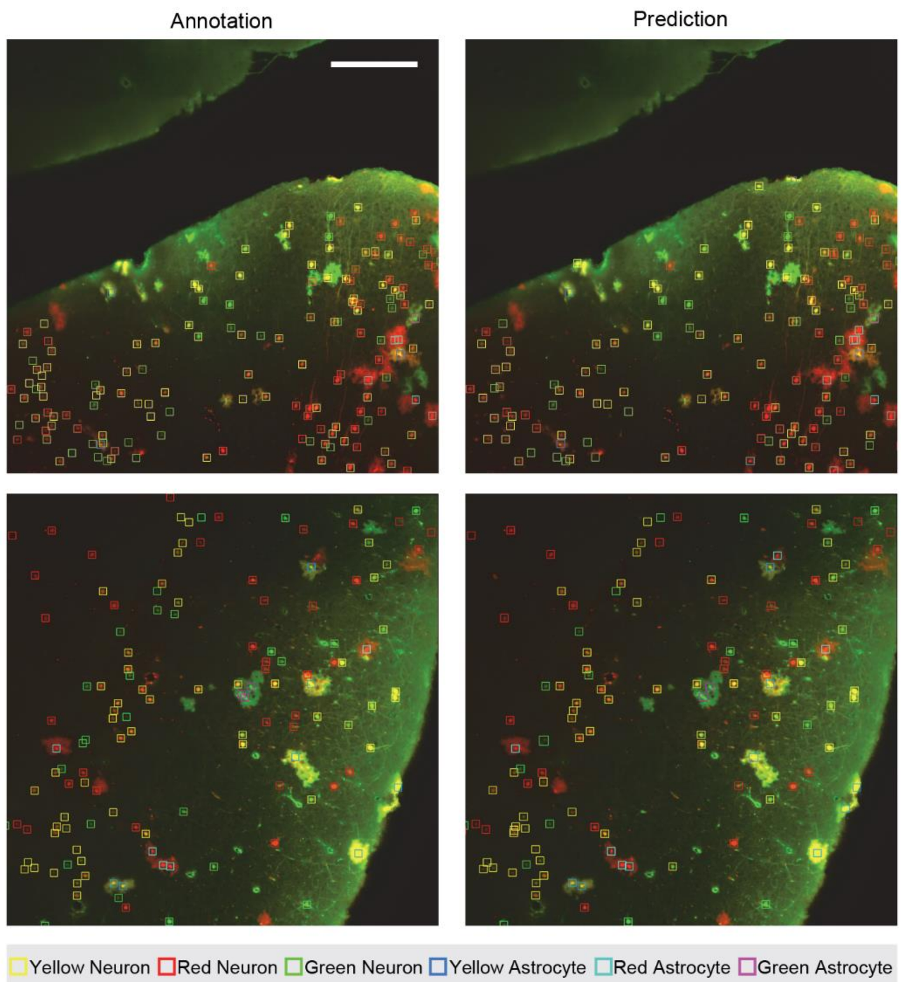


Figure S2. Comparison of manual annotation versus automated prediction on field-of-view images. Related to Figure 2. Scale bar, 200 µm.

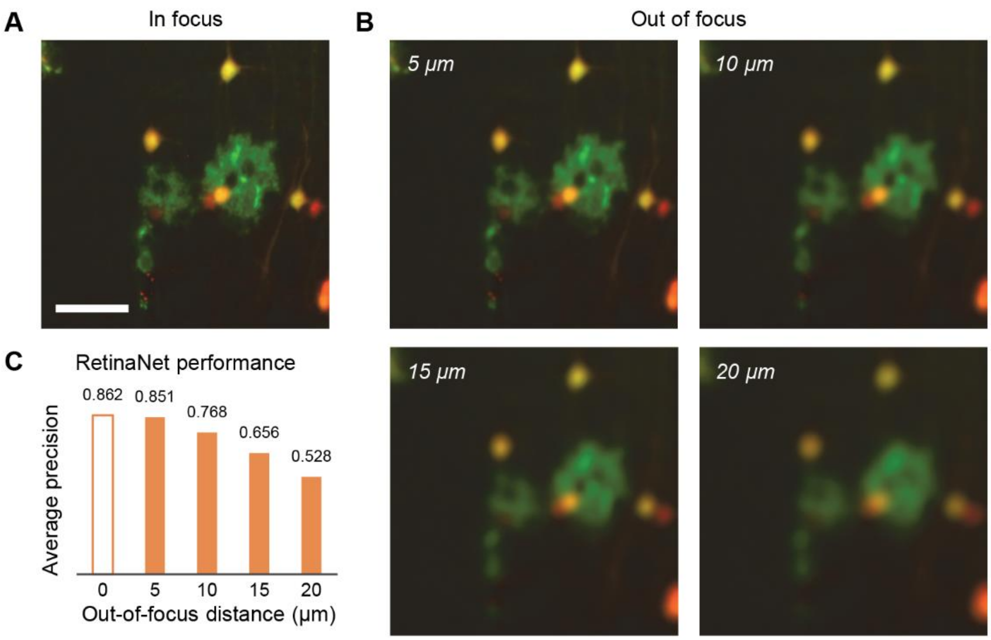


Figure S3. Comparison of RetinaNet performance with and without defocus. Related to Figure 3. (A) An example of an in-focus image. Scale bar, 50 μm. **(B)** The corresponding synthetic out-of-focus images at different defocus distances from the focal plane. **(C)** Average precision results of the RetinaNet detection on in-focus and synthetic out-of-focus test data. RetinaNet performed well when the defocus distance from the focal plane was within 5 μm.

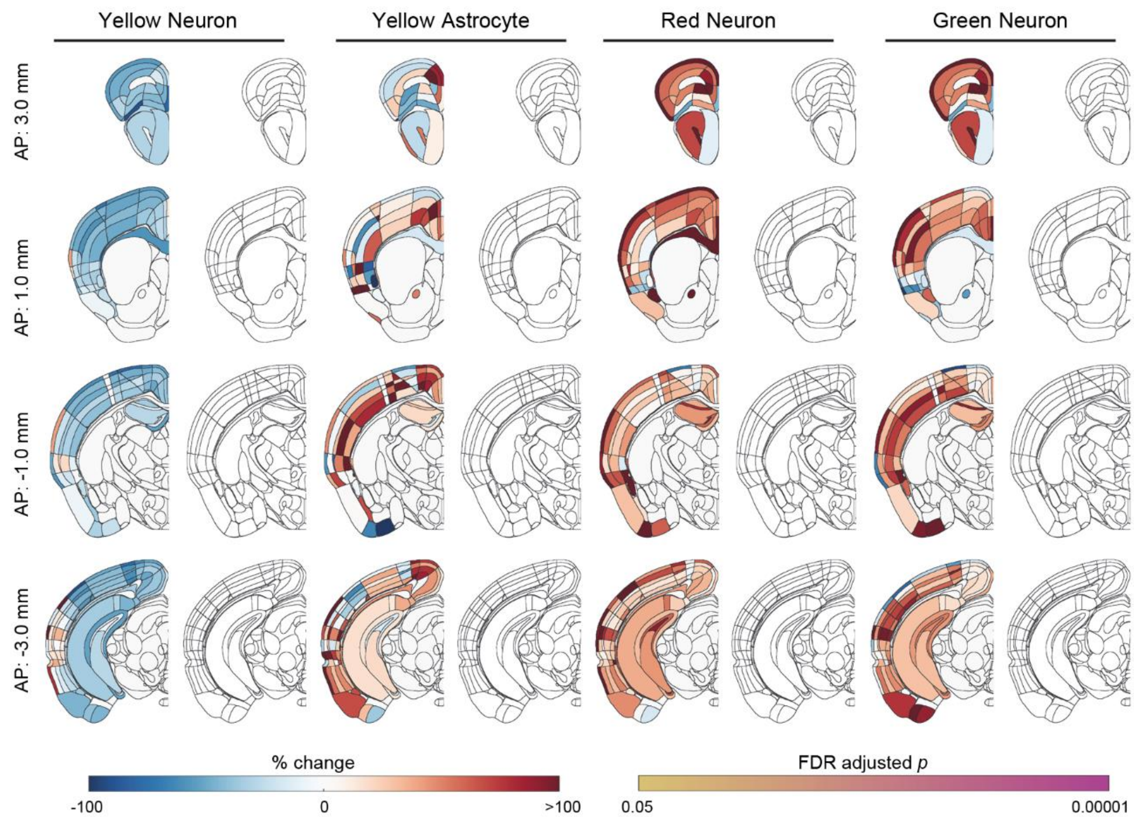


Figure S4. Region-wise analysis of average cell densities between *Emx:MADM:+/+* and *Emx:MADM:F/+* datasets. Related to Figure 4. Left panels: percentage change in average cell densities of *Emx:MADM:F/+* brain hemispheres compared to *Emx:MADM:+/+* brain hemispheres. Right panels: adjusted *p*-values (number of animals = 3). AP, anteroposterior distance taken from the bregma; FDR, false discovery rate.

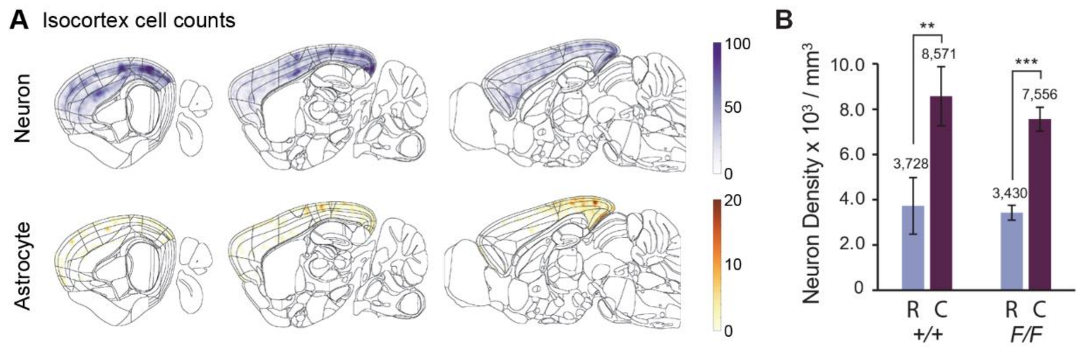


Figure S5. Variance in neuron distribution between rostral and caudal cortices. Related to Figure 6. (A) Average voxel-wise cell density maps of *Emx:MADM:F/F* brain hemispheres (number of animals = 3). **(B)** Significant difference in MADM-labeled neuron densities was observed between rostral and caudal regions in both *Emx:MADM:+/+* and *Emx:MADM:F/F* brain hemispheres (bar chart: mean ± SD, number of animals = 3; **, $p < 0.01$; ***, $p < 0.001$).

2 Supplemental tables

Table S1. Precision of cell detection at confidence threshold of 0.5. Related to Figure 2.

	Neuron	Astrocyte	Average
Yellow	0.863	0.787	0.825
Green	0.863	0.894	0.879
Red	0.907	0.774	0.841
Average	**0.878**	**0.818**	**0.848**

Table S2. Recall of cell detection at confidence threshold of 0.5. Related to Figure 2.

	Neuron	Astrocyte	Average
Yellow	0.874	0.738	0.806
Green	0.709	0.766	0.738
Red	0.735	0.800	0.768
Average	**0.773**	**0.768**	**0.771**

Table S3. F-score of cell detection at confidence threshold of 0.5. Related to Figure 2.

	Neuron	Astrocyte	Average
Yellow	0.868	0.762	0.815
Green	0.778	0.825	0.802
Red	0.812	0.787	0.800
Average	**0.819**	**0.791**	**0.806**

Table S4. 31 regions show significant increases of red MADM astrocyte. Related to Figure 4.

Name	% Change	Adjusted *p*-value
Cortical amygdalar area	2379.618	0.0027
Cortical amygdalar area posterior part	2530.341	0.0027
Primary motor area Layer 1	659.7982	0.0038
Entorhinal area lateral part layer 3	418.6971	0.0038
Corpus callosum	469.5149	0.0074
Dentate gyrus granule cell layer	1265.942	0.0107
Orbital area	328.3208	0.0121
Retrosplenial area lateral agranular part layer 2/3	247.3442	0.0182
Genu of corpus callosum	438.9664	0.0221
Entorhinal area medial part dorsal zone layer 5	445.1044	0.0240
Retrosplenial area dorsal part	515.8460	0.0261
Anteromedial visual area layer 5	627.4307	0.0265
Agranular insular area	501.1534	0.0274
Supplemental somatosensory area layer 6b	306.5585	0.0309
Ventral auditory area layer 5	1369.305	0.0419
Anteromedial visual area	338.6888	0.0419
Posteromedial visual area layer 1	937.5416	0.0419
Retrosplenial area	425.3429	0.0419
Retrosplenial area dorsal part layer 1	372.9179	0.0419
Retrosplenial area ventral part	395.9375	0.0419
Entorhinal area lateral part	442.4262	0.0419
Entorhinal area lateral part layer 2	383.3483	0.0419
Lateral forebrain bundle system	435.2953	0.0419
Retrosplenial area ventral part layer 5	504.5898	0.0438
Temporal association areas layer 6a	825.1289	0.0450
Entorhinal area medial part dorsal zone layer 3	308.5111	0.0450
Primary somatosensory area upper limb layer 1	328.5355	0.0454
Triangular nucleus of septum	538.6442	0.0454
Olfactory areas	250.8286	0.0475
Anterior cingulate area dorsal part	599.8512	0.0499
Infralimbic area layer 6a	609.0158	0.0499

Table S5. Regions showing significant differences in hierarchical analysis. Related to Figure 5.

Level	Name
1	Corpus callosum
2	Genu of corpus callosum
3	-
4	Cortical amygdalar area Orbital area Agranular insular area Retrosplenial area Triangular nucleus of septum
5	Cortical amygdalar area posterior part Secondary motor area Gustatory areas layer 5 Visceral area layer 4 Anteromedial visual area Posterolateral visual area Anterior cingulate area dorsal part Infralimbic area layer 5 Infralimbic area layer 6a Infralimbic area layer 6b Retrosplenial area lateral agranular part Retrosplenial area dorsal part Retrosplenial area ventral part Temporal association areas layer 5 Temporal association areas layer 6a Dentate gyrus Entorhinal area Prosubiculum
6	Primary motor area Layer 5 Secondary motor area layer 1 Secondary motor area layer 2/3 Secondary motor area layer 5 Supplemental somatosensory area layer 6b Dorsal auditory area layer 4 Ventral auditory area layer 5 Ventral auditory area layer 6b Anteromedial visual area layer 1 Anteromedial visual area layer 5 Anteromedial visual area layer 6a Anteromedial visual area layer 6b Lateral visual area layer 4

Table S5. Regions showing significant differences in hierarchical analysis. Related to Figure 5. (continued)

	Primary visual area layer 1 Posterolateral visual area layer 2/3 Posterolateral visual area layer 4 Posteromedial visual area layer 1 Posteromedial visual area layer 2/3 Anterior cingulate area dorsal part layer 5 Anterior cingulate area dorsal part layer 6a Agranular insular area dorsal part layer 2/3 Retrosplenial area lateral agranular part layer 1 Retrosplenial area lateral agranular part layer 2/3 Retrosplenial area lateral agranular part layer 6b Retrosplenial area dorsal part layer 1 Retrosplenial area dorsal part layer 2/3 Retrosplenial area dorsal part layer 5 Retrosplenial area ventral part layer 5 Cortical amygdalar area posterior part medial zone Dentate gyrus molecular layer Dentate gyrus polymorph layer Entorhinal area lateral part Entorhinal area medial part dorsal zoneGenu of corpus callosum

Appendix C: CHAPTER 4 Supplemental information

1 Supplemental figures

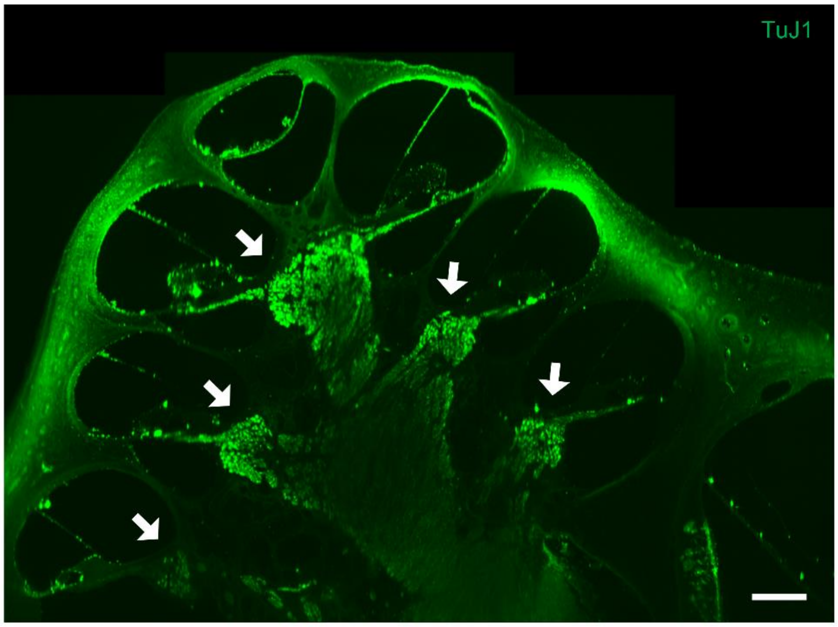


Figure S1. A virtual cross section of pig cochlea shows locations of spiral ganglion neurons. The virtual slice was from a 2-week-old pig cochlea stained for TuJ1, imaged using a light sheet fluorescence microscope. Spiral ganglion neurons (arrow) are located in the modiolus within the cochlea. Scale bar, 500 µm.

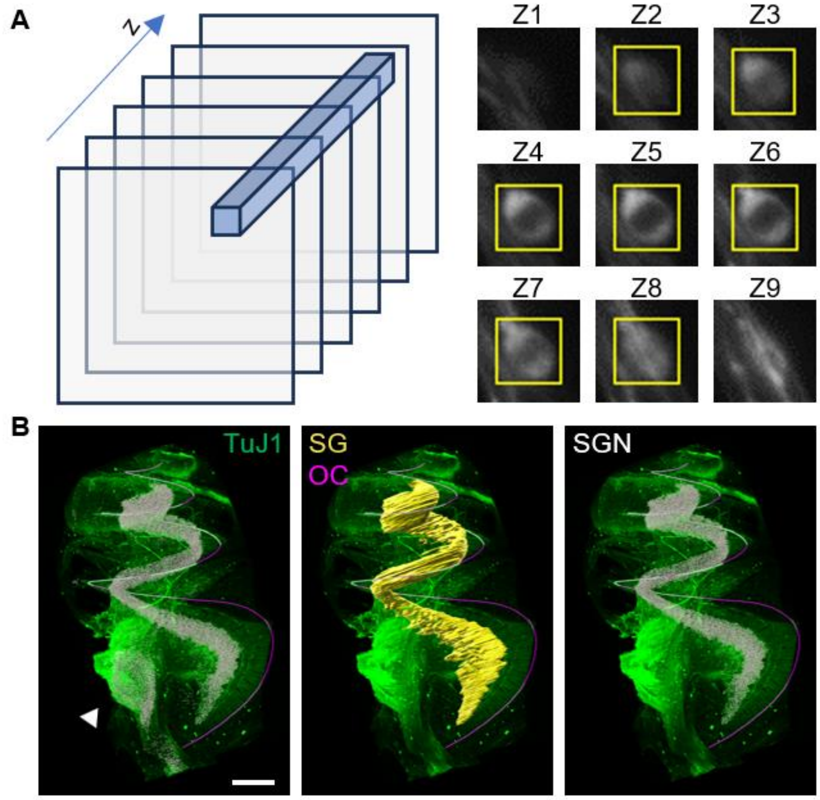


Figure S2. (A) A spiral ganglion neuron (SGN) appeared in multiple z-planes after light sheet imaging and was detected by the cell detection network repeatedly (yellow box). The dataset was acquired at a z step size of 2 μm. **(B)** Neurons adjacent to the nerve bundle outside the modiolus were also detected by the network (arrowhead). Therefore, a segmentation mask was applied to exclude detections outside the spiral ganglion (SG, yellow). OC, organ of Corti. Scale bar, 400 μm.

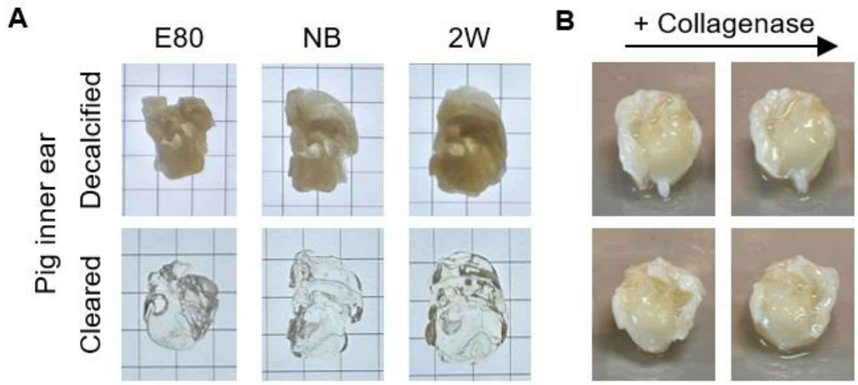


Figure S3. (A) Pig inner ears at different ages before and after tissue clearing. E80, embryonic day 80; NB, newborn; 2W, 2 weeks. The samples were imaged in the air. Grid size, 5 mm. **(B)** Two 6–8-week-old pig's inner ears before and after collagenase treatment. Collagenase treatment didn't change the sample sizes and structures dramatically.

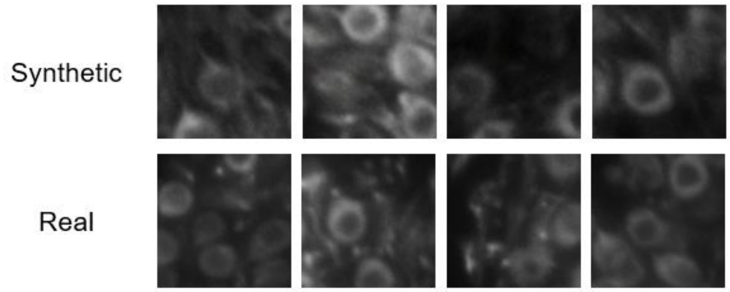


Figure S4. Synthetic spiral ganglion neuron images generated by a diffusion model. The model was trained on the gerbil light sheet dataset of spiral ganglion neurons with TuJ1 staining.