Quantifying cellular parameters across the murine brain: New practices for integrating and analysing neuroscience data using 3D brain atlases

Ingvild Elise Bjerke

A thesis submitted
to the University of Oslo
for the degree of

Philosophiae Doctor (PhD)

2020

Principal supervisor: Trygve B. Leergaard

Co-supervisors: Maja A. Puchades and Jan G. Bjaalie

Division of Anatomy
Department of Molecular Medicine
Institute of Basic Medical Sciences
University of Oslo

Series of dissertations submitted to the
Faculty of Medicine, University of Oslo

ISBN 978-82-8377-792-5

Cover: Hanne Baadsgaard Utigard.
Print production: Reprosentralen, University of Oslo.

Contents

Acknowledgements

The work presented in this thesis was performed between 2017 and 2020 at the Neural Systems Laboratory, Institute of Basic Medical Sciences, University of Oslo. The project was funded by the European Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreement No. 720270 (Human Brain Project SGA1), Specific Grant Agreement No. 785907 (Human Brain Project SGA2), Specific Grant Agreement No. 945539 (Human Brain Project SGA3), and the University of Oslo, Norway, with additional support from The Research Council of Norway under Grant Agreement No. 269774 (INCF Norwegian Node).

First and foremost, I want to thank my primary supervisor Trygve Leergaard. He has shared his passion for and knowledge of science and writing in a way that made it both educational and inspiring to work on the project. Despite his very busy schedule, he always managed to set aside time for valuable meetings. I have especially appreciated all the thought-provoking discussions, but also his enthusiastic and positive spirit and attention to detail.

I am also very grateful to my two co-supervisors, Maja Puchades and Jan Bjaalie, for their advice and support throughout the project. I want to thank Maja for always providing useful insights and advice regarding experimental work and analytic procedures. Jan has shared of his knowledge of neuroinformatics projects and resources, always in a constructive and inspirational way.

My thanks also go to all my wonderful colleagues in the Neural Systems laboratory. Large parts of this thesis was written in relative isolation due to the COVID-19 pandemic. Although not an optimal situation, it made me appreciate even more the close collaborations and frequent discussions we normally have. Special thanks goes to my office partners, Heidi Kleven and Martin Øvsthus. I am deeply grateful for all the discussions we have had, especially the virtual ones during the long weeks of home offices. Also, I am thankful for their helpful comments on figures (which saved me from a stroke several times when trying to understand Photoshop or Illustrator!), but not least for providing such a positive atmosphere

and for making me work out once in a while. Furthermore, I am grateful to Jala Imad for our collaboration and discussions regarding the basal ganglia delineations in the Waxholm Space rat brain atlas; these new delineations were critically important for creating a database of literature derived data from the basal ganglia.

I would also like to thank all the co-authors on the publications and data sets presented in this thesis. In particular, I am thankful to Ulrike Schlegel for her assistance and advice when sharing data sets through the EBRAINS infrastructure. Her help with organising and curating all the spatial co-registration datasets was of critical importance to get them shared. I also want to thank Sharon Yates, whose work on the QUINT workflow and insights regarding analytic procedures were critical for the last part of this project. Furthermore, I would like to thank Arthur Laja for acquiring the beautifully stained parvalbumin sections used in the last part of this project and for his hard work on organising these data for sharing.

Importantly, I would like to thank Grazyna Babinska and Hong Qu for their expert technical help with staining and scanning sections, respectively. I am also very grateful to the developers at the Neural Systems laboratory, Nicolaas Groeneboom, Gergely Csucs and Dmitri Darine, for making so many fantastic applications and always providing advice on technical challenges.

I am truly thankful to the friends and family members who have supported me throughout the years working on this thesis. Special thanks to Madelaine, Siren and Marte for giving me an excuse to have a night off on a regular basis, which I especially appreciated during the last year. A big thanks also to Anna, for always listening to my long winded stories about the ups and downs of the project and for seemingly being interested in hearing about it as well. Lastly, I want to thank my parents and my two sisters, for always being there and for giving me places to run away and write whenever I needed it.

Oslo, August 2020

Ingvild E. Bjerke

Norsk sammendrag

Det investeres mye i forskning for å forstå hjernens normale funksjon og de mange sykdommene som kan ramme den. Stadig mer avanserte teknikker anvendes og antallet nevrovitenskapelige publikasjoner øker jevnt. Men, det er krevende å få oversikt over den tilgjengelige informasjonen. Fordi de fleste studier er utført med forskjellige metoder og rapportert på ulike (og ofte mangelfulle) måter kan resultatene være vanskelige å sammenligne. Dette gjør det vanskelig å svare på tilsynelatende enkle spørsmål, som for eksempel "hvor mange celler finnes i dette området av hjernen?". For å besvare grunnleggende problemstillinger og legge til rette for fremgang i feltet er det nødvendig at informasjon fra ulike studier kan sammenstilles, eller integreres, i en felles kontekst.

Integrasjon av informasjon fra mange studier krever standardisering. Nevrovitenskapelige forskningsdata omfatter mange typer data, som varierer både i tid, rom og format. Noen data viser fysiologiske fenomener med millisekundspresisjon, andre representerer prosesser som skjer over måneder og år. Nevrovitenskapelig bildemateriale spenner fra radiologiske fremstillinger av struktur og funksjon i hele hjerner, til mikroskopiske bilder av celler, synapser og molekyler på mikro- og nanometernivå. En fellesnevner for slike data er imidlertid at de har sitt fysiske opphav i en anatomisk posisjon i hjernen. Hjernen er organisert i ulike systemer, der flere områder med ulike egenskaper er forbundet i større nettverk. Enhver observasjon fra hjernen må derfor tolkes ut fra *hvor* den er innhentet. Informasjon om anatomisk posisjon er dermed avgjørende for å tolke, sammenligne, og til slutt integrere data. For å avgjøre hvor i hjernen en observasjon stammer fra, benytter forskere ofte hjerneatlas. Tradisjonelt består disse av diagrammer som er basert på todimensjonale snitt gjennom hjernen, presentert i bokform. Selv om slike atlas gir god oversikt over hjernens strukturer, gir de begrensede muligheter til å gjøre sammenligninger med nytt materiale, som gjerne er presentert på en annen måte. Nye digitale, tredimensjonale hjerneatlas gir bedre muligheter for å visualisere og sammenligne anatomisk informasjon.

Hovedmålet med arbeidet som presenteres i denne avhandlingen var å etablere og teste nye standarder og metoder for å samle, integrere og dele nye og

tidligere publiserte nevrovitenskapelige data gjennom bruk av digitale hjerneatlas. Fokuset ble spesielt rettet mot analyse og deling av kvantitative cellulære data.

I første del av arbeidet spesifiserte vi minstekrav til dokumentasjon av anatomiske posisjoner i hjernen, og viste hvordan nye metoder for bruk av digitale hjerneatlas kan utnyttes til å kartlegge og dokumentere hvor i hjernen data kommer fra på en mer standardisert måte. Vi undersøkte litteraturen for å finne ut hvordan anatomiske posisjoner vanligvis dokumenteres. Dette viste at rapporteringen av anatomiske posisjoner overraskende ofte er upresis eller mangelfull, og at det er varierende praksis for å dokumentere og rapportere anatomisk posisjon i ulike typer studier. Våre funn viste at det med relativt enkle grep er mulig å forbedre rapporteringen av anatomiske posisjoner, slik at det blir lettere å tolke informasjonen. Vi gav deretter eksempler på hvordan nye metoder for å knytte data til hjerneatlas kan brukes til å integrere ulike typer data.

I andre del av arbeidet fokuserte vi på kvantitative anatomiske data fra basalganglieområdene hos rotte og mus. Etter å ha undersøkt hvilke parametere det finnes informasjon om i litteraturen, valgte vi å fokusere på kvantitativ informasjon om celler, synapser og dendrittgrener. Vi undersøkte hvordan disse kan samles og integreres, og vurderte i hvilken grad rapporterte tall er sammenlignbare på tvers av studier. Vi laget en relasjonell database der vi samlet tall fra publiserte artikler, sammen med informasjon om dyr, metoder, analyser og anatomiske posisjoner. Fordi anatomiske posisjoner rapporteres på mange ulike måter, og fordi forskere ikke alltid er enige om hvordan områder bør navngis, etablerte vi metoder for å oversette de mange anatomiske begrepene som brukes i litteraturen til en felles, standardisert terminologi. Vi gjorde systematiske litteratursøk for å finne kilder, og samlet over 1000 rapporterte tall i databasen. Hele databasen ble publisert som et offentlig tilgjengelig datasett via den nye europeiske forskningsinfrastrukturen EBRAINS. Ved gjennomgang av all informasjonen i databasen viste vi at de rapporterte tallene for tilsvarende parametere varierer mye på tvers av studier, og at de fleste celletyper sjelden blir undersøkt og kvantifisert. Også her fant vi at dokumentasjonen som gis i artikler ofte er mangelfull, og for eksempel mangler viktig informasjon om hvordan dataene er produsert og analysert.

I siste del av arbeidet gjennomførte vi en kvantitativ analyse av to spesifikke celletyper gjennom hele hjernen hos rotte og mus. Immunohistokjemiske metoder ble brukt for å visualisere celler positive for parvalbumin- og calbindin-protein. Vi knyttet alle dataene til digitale hjerneatlas, og brukte en semi-automatisk, maskinlæringsbasert metode for å gjenkjenne cellene i bildematerialet. Deretter kvantifiserte vi cellene for hvert område av hjernen. Vår analyse viste at de to celletypene har en komplementær distribusjon i musehjernen,

med parvalbumin-positive celler som den primære typen i corticale og hippocampale områder, mens calbindin-positive celler er vanligere i limbiske, thalamiske og hypothalamiske områder. Både originaldata, prosesserte data, og analyseresultater fra dette studiet ble gjort tilgjengelig via EBRAINS i henhold til FAIR-prinsippene, slik at andre kan gjenbruke og bygge videre på resultatene. I tillegg til de biologiske funnene, viser studiet hvordan kvantitative anatomiske data fra hjernen effektivt kan analyseres, rapporteres og deles på en åpen måte som legger opp til gjenbruk.

Arbeidet som inngår i avhandlingen bidrar først og fremst med nye standarder og metoder for å rapportere, analysere og dele forskningsdata innen nevrovitenskap. I tillegg har vi vist hvordan store mengder kvantitative data kan kartlegges systematisk, og integreres i en database. Dette arbeidet viser hvordan nevrovitenskapelige data kan samles og integreres på en ny måte som vil gjøre det lettere å tolke og å gjenbruke andres resultater. Vi er overbevist om at dette vil være essensielt for å bringe feltet videre mot å løse de store spørsmålene om hjernen.

List of papers

- **Bjerke IE**, Øvsthus M, Andersson KA, Blixhavn CH, Kleven H, Yates SC, Puchades MA, Bjaalie JG & Leergaard TB **Navigating the Murine Brain: Toward Best Practices for Determining and Documenting Neuroanatomical Locations in Experimental Studies** Frontiers in Neuroanatomy 12 (2018) DOI: 10.3389/fnana.2018.00082

- **Bjerke IE**, Puchades M, Bjaalie JG & Leergaard TB **Database of literature derived cellular measurements from the murine basal ganglia** Scientific Data (2020) DOI: 10.1038/s41597-020-0550-3

- **Bjerke IE**, Yates SC, Laja A, Witter MP, Puchades M, Bjaalie JG & Leergaard TB **Densities and numbers of calbindin and parvalbumin positive neurons across the rat and mouse brain: Complementary patterns and cross-species comparison** Submitted manuscript.

List of data sets

- Bjerke IE, Puchades M, Bjaalie, JG & Leergaard TB **Database of quantitative cellular and subcellular morphological properties from rat and mouse basal ganglia** Human Brain Project Neuroinformatics Platform (2019) DOI: 10.25493/DYXZ-76U

- Bjerke IE, Puchades M, Bjaalie JG & Leergaard TB **Comparability of basal ganglia delineations across different mouse brain atlases** Human Brain Project Neuroinformatics Platform (2019) DOI: 10.25493/MWAS-3S6

- Bjerke IE, Puchades M, Bjaalie JG & Leergaard TB **Comparability of basal ganglia delineations across different rat brain atlases** Human Brain Project Neuroinformatics Platform (2019) DOI: 10.25493/D2M9-BSK

- Bjerke IE, Schlegel U, Puchades M, Bjaalie, JG & Leergaard, TB **Franklin & Paxinos' "The Mouse Brain in Stereotaxic Coordinates" (3rd edition) spatially registered to the Allen mouse brain Common Coordinate Framework** Human Brain Project Neuroinformatics Platform (2019) DOI:10.25493/1BT9-YYD

- Bjerke IE, Schlegel U, Puchades M, Bjaalie, JG & Leergaard, TB **Franklin & Paxinos' "The Mouse Brain in Stereotaxic Coordinates" (4th edition) spatially registered to the Allen Mouse Common Coordinate Framework** Human Brain Project Neuroinformatics Platform (2019) DOI:10.25493/WFCZ-FSN

- Bjerke IE, Schlegel U, Puchades M, Bjaalie, JG & Leergaard, TB **Paxinos & Franklin's "The Mouse Brain in Stereotaxic Coordinates" (2nd edition) spatially registered to the Allen Mouse Common Coordinate Framework** Human Brain Project Neuroinformatics Platform (2019) DOI: 10.25493/BTKK-CRY

- Bjerke IE, Schlegel U, Puchades M, Bjaalie, JG & Leergaard, TB **Paxinos & Watson's "The Rat Brain in Stereotaxic Coordinates" (1st**

edition) spatially registered to the Waxholm Space atlas of the rat brain Human Brain Project Neuroinformatics Platform (2019) DOI: 10.25493/YRKH-626

- Bjerke IE, Schlegel U, Puchades M, Bjaalie, JG & Leergaard, TB **Paxinos & Watson's "The Rat Brain in Stereotaxic Coordinates" (3rd edition) spatially registered to the Waxholm Space atlas of the rat brain** Human Brain Project Neuroinformatics Platform (2019) DOI: 10.25493/KNB2-GMN

- Bjerke IE, Schlegel U, Puchades M, Bjaalie, JG & Leergaard, TB **Paxinos & Watson's "The Rat Brain in Stereotaxic Coordinates" (4th edition) spatially registered to the Waxholm Space atlas of the rat brain** Human Brain Project Neuroinformatics Platform (2019) DOI: 10.25493/W3R1-R4A

- Bjerke IE, Schlegel U, Puchades M, Bjaalie, JG & Leergaard, TB **Paxinos & Watson's "The Rat Brain in Stereotaxic Coordinates" (5th edition) spatially registered to the Waxholm Space atlas of the rat brain** Human Brain Project Neuroinformatics Platform (2019) DOI: 10.25493/KQ5K-S0D

- Bjerke IE, Schlegel U, Puchades M, Bjaalie, JG & Leergaard, TB **Paxinos & Watson's "The Rat Brain in Stereotaxic Coordinates" (6th edition) spatially registered to the Waxholm Space atlas of the rat brain** Human Brain Project Neuroinformatics Platform (2019) DOI: 10.25493/XQ8J-TNE

- Bjerke IE, Schlegel U, Puchades M, Bjaalie, JG & Leergaard, TB **Paxinos & Watson's "The Rat Brain in Stereotaxic Coordinates" (7th edition) spatially registered to the Waxholm Space atlas of the rat brain** Human Brain Project Neuroinformatics Platform (2019) DOI: 10.25493/APWV-37H

- Bjerke IE, Schlegel U, Puchades M, Bjaalie, JG & Leergaard, TB **Paxinos and colleagues' "MRI/DTI Atlas of the Rat Brain" (1st edition) spatially registered to the Waxholm Space atlas of the rat brain** Human Brain Project Neuroinformatics Platform (2019) DOI: 10.25493/9BHD-WDP

- Bjerke IE, Schlegel U, Puchades M, Bjaalie, JG & Leergaard, TB **Swanson's "Brain Maps: Structure of the Rat Brain" (1st edition) spatially registered to the Waxholm Space atlas of the rat brain** Human Brain Project Neuroinformatics Platform (2019) DOI: 10.25493/ZB03-H5G

- Bjerke IE, Schlegel U, Puchades M, Bjaalie, JG & Leergaard, TB **Swanson's "Brain Maps: Structure of the Rat Brain" (2nd edition) spatially registered to the Waxholm Space atlas of the rat brain** Human Brain Project Neuroinformatics Platform (2019) DOI: 10.25493/EEQA-9RM

- Bjerke IE, Schlegel U, Puchades M, Bjaalie, JG & Leergaard, TB **Swanson's "Brain Maps: Structure of the Rat Brain" (3rd edition) spatially registered to the Waxholm Space atlas of the rat brain** Human Brain Project Neuroinformatics Platform (2019) DOI: 10.25493/ZFXB-23F

- Bjerke IE, Schlegel U, Puchades M, Bjaalie, JG & Leergaard, TB **Swanson's "Brain Maps: Structure of the Rat Brain" (4th edition) spatially registered to the Waxholm Space atlas of the rat brain** Human Brain Project Neuroinformatics Platform (2019) DOI: 10.25493/486N-966

- Bjerke IE & Leergaard, TB **Distribution of calbindin positive neurons in the normal adult mouse brain** EBRAINS (2020) DOI: 10.25493/KHNT-KV8

- Laja A, Bjerke IE, Leergaard TB & Witter, MP **Distribution of parvalbumin-positive interneurons in the normal adult mouse brain** EBRAINS (2020) DOI: 10.25493/BXGX-WM4

- Laja A, Bjerke IE, Leergaard TB & Witter, MP **Distribution of parvalbumin-positive interneurons in the normal adult rat brain** EBRAINS (2020) DOI: 10.25493/8KCQ-3C7

- Bjerke IE, Yates SC, Puchades M, Bjaalie JG & Leergaard, TB **Brain-wide quantitative data on calbindin positive neurons in the mouse** EBRAINS (2020) DOI: 10.25493/TT2Y-23N

- Bjerke IE, Yates SC, Puchades M, Bjaalie JG & Leergaard, TB **Brain-wide quantitative data on parvalbumin positive neurons in the rat** EBRAINS (2020) DOI: 10.25493/KR92-C33

- Bjerke IE, Yates SC, Puchades M, Bjaalie JG & Leergaard, TB **Brain-wide quantitative data on parvalbumin positive neurons in the mouse** EBRAINS (2020) DOI: 10.25493/BT8X-FN9

Key terms used in the thesis

Primary data. Also often referred to as raw data. Defined here as the (physical or digital) result of an experiment that is considered, for any given method, to be the first generation of data. Digital primary data might be processed, as different methods require various levels of processing for primary data to be considered legible.

Derived data. A data element derived from other data elements using a mathematical, logical, or other type of transformation, e.g. arithmetic formula, composition, aggregation. *Definition adopted from OECD glossary of statistical terms, https://stats.oecd.org/glossary/detail.asp?ID=5130.*

Data. Collective term used here to refer to both primary and derived data.

Metadata. Data that provides information about other data. *Definition adopted from Merriam-Webster dictionary, www.merriam-webster.com/dictionary/metadata.*

- **Anatomical metadata.** Defined here as terms and / or coordinates describing anatomical locations - in the context of this thesis, within the brain.

Data integration. Data integration involves combining data from several disparate sources, which are stored using various technologies and provide a unified view of the data. *Definition adopted from Data Integration Information, www.dataintegration.info/data-integration.*

- **Location integration.** Defined here as data integration through the use of anatomical metadata.

Atlas registration. Defined here as the process of defining anatomical location of data in a brain atlas, semantically (using anatomical terms) or based on spatial coordinates.

- **Spatial atlas registration.** Defined here as the process of defining anatomical location of data in an atlas by use of the atlas coordinates.
- **Semantic atlas registration.** Defined here as the process of defining anatomical location of data in an atlas by use of the atlas terminology.

Preface

So, oft in theologic wars
The disputants, I ween,
Rail on in utter ignorance
Of what each other mean;
And prate about an Elephant
Not one of them has seen!
–John Godfrey Saxe

In an ancient parable, we learn about a group of blind men encountering an elephant for the first time. Wishing to know what this creature is, they all begin exploring the elephant: touching, prodding, and stroking until they are certain they must know the nature of the thing in front of them. Each man, however, touches a different part of the animal. When they gather to discuss the concept of an elephant, the man who had touched only the tusks has a very different opinion than he who had felt the leathery skin or the tail. Blind as they all are to the bigger picture, the discussion immediately becomes heated. In some versions of the story they come to distrust each others honesty and reliability. The central challenge is that each man is presenting his data as if it was knowledge. In information science, data is often viewed as the lowest level in a hierarchy via information, to knowledge, and - ultimately - to wisdom (or understanding; [1]). A prerequisite for converting data to knowledge in any scientific field is to combine data from various sources. The story of the blind men and the elephant tells us that as humans, we tend to represent our partial experiences as the whole truth. We also tend to disregard the experiences relayed by others. These insights have direct implications for science: it tells us that our observations must be considered partial experiences that needs to be put in context of other researchers' observations in order to become meaningful.

Combining data from multiple sources is particularly challenging in neuroscience. The brain is our most complex and important organ, involved in anything and everything we do: when we brush our teeth in the morning, when

we learn to ride a bike, when we smile politely at a remote contact or when we care for our loved ones. Neuroscience research spans scales in space and time - from tiny proteins and synapses to macroscopic creases and folds, from millisecond events to ageing processes occurring over years and even decades. With advances in technologies, the questions being asked in neuroscience research are becoming increasingly sophisticated and datasets are becoming larger and more complex to interpret. Yet, findings are still primarily conveyed through the research article format, with selected data and example images presented. It has been noted for some time that this traditional format will not suffice to meet the challenges facing neuroscientists in the 21st century [2], and that data should be themselves be considered first class research objects [3, 4]. Fostering a culture of sharing data has been recognized as an important prerequisite for meeting these challenges and to further our understanding of the brain [5–7]. To facilitate sophisticated search, interpretation and re-use of data, they must also be combined and presented in a unified view with other data, here referred to as *data integration*. Had the blind men from the parable discussed above integrated their data, they would quickly have realised that their observations differed depending on which part of the elephant they touched.

Building the infrastructure, standards, and software to enable sharing, integration and analysis of experimental neuroscience data is an active pursuit within the field of neuroinformatics [8, 9], where large-scale initiatives have been undertaken to achieve progress. The NIH-funded Human Brain Project and Biomedical Informatics Research Network, launched in the 1990's and early 2000's, respectively, were pioneering projects in this regard [2]. Over the last decade, several large-scale projects have been launched across the globe: examples include the Brain Initiative in the US [10], the Canadian brain-CODE project [11], and the brain/MINDS project in Japan [12]. In Europe, the Human Brain Project aims to build a research infrastructure enabling approaches across neuroscience, computing, informatics and brain-inspired technology [13], while the Blue Brain Project is dedicated to build detailed reconstructions and computational models of the murine and human brain [14]. Hopefully, by providing the urgently needed tools to cope with the amounts of data being produced, such projects will allow our understanding of the brain to progress more rapidly.

Chapter 1

Background

1.1 Introduction

The importance of the brain in every aspect of our life is mirrored in the devastating impact of its diseases. Brain diseases account for 23 % of the loss of healthy years and 50 % of years of life lived with disability, thus being the single greatest cause of disease burden [15]. Substantial efforts have been made to advance our understanding our the brain, but with the extreme growth rate of scientific output [16] (referred to as the 'data deluge'), it is all but impossible to navigate the body of research to overview even narrow fields of interest. Researchers struggle to reconcile different or contradictory findings, and results often cannot be reproduced [17]. Thus, progress towards improving our understanding of the brain has been slower than anticipated, and therapies for some of the most debilitating diseases of our time are still lacking. To tackle the complexity of the brain and improve our understanding across different levels of research, coordinated efforts to acquire and share data will be essential. In particular, there is a need to combine data from multiple sources, here referred to as *data integration*.

One could consider several levels of data integration. In a way, descriptive combinations of two or more data sets, typically provided in reviews, constitute a form of data integration. At a more technical level, data integration involves normalising data and standardising the information about them so that they can be compared and presented in a unified view [2, 18] - this level of data integration is primarily the focus of this thesis. The deepest level of data integration in neuroscience would be combining them in novel theories or computational models of the brain [19, 20]. The databases and tools needed to support large-scale acquisition, sharing, analysis and integration of data are developed through neuroinformatics, an interdisciplinary field that combines neuroscience and information science [8].

An important step towards data integration is annotation of data with extensive information about their properties, acquisition, conditions of use, etcetera [21]. Such "data about data" are referred to as metadata. In principle, a research article should contain all the metadata necessary for an independent researcher to interpret and replicate a study, but in many cases, metadata reported in publications are insufficient for researchers to fully interpret the findings [22–24]. On the other hand, when metadata are highly structured and well-defined, they represent a powerful means to integrate, query and filter data [25, 26]. Metadata are thus relevant both for documentation of findings and as a tool for indexing data. For both purposes, there is a growing need for standards specifying which metadata need to be reported and - importantly - how to report them (e.g. through standard terminologies). A challenge is that concepts about neuroscience metadata are

constantly evolving, with naming schemes for brain regions [27] and cell types [28, 29] being particularly debated. Researchers may find it difficult to adhere to a standard if they do not fully agree with its definitions, and may simply use the terms they prefer or even arbitrary terms ("cell type a", "cell type b"), which has led to a confusing landscape of terms. Thus, efforts are needed not only to create metadata standards, but also to promote their adoption by the community.

Recognising the need to gather neuroscience data and present it in a unified format, several valuable repositories containing various types of primary or derived data have been built through neuroinformatics projects. For example, the Allen Institute for Brain Science provides access to large amounts of primary data [30], including *in situ* hybridisation [31], connectivity [32], and RNA-sequencing material [33]. Several other repositories exist as well, with access to computational models [34], neuronal reconstructions [35, 36], neural connectivity data [37–40], microscopy image data [41], or all of the above [13]. In parallel with the efforts to share primary data, collecting and curating published information is important to make use of the collective knowledge in the field. Significant effort is required to extract and standardise data and metadata from research papers. Thus, databases containing derived data from the literature are especially valuable in that they provide easy overview of information that would otherwise be distributed across text, figures and tables of papers, in terminologies that depend on the context or the particular preference of the authors. Such derived data may be used to identify knowledge gaps, constrain computational models, or compare new data to existing information. Much of the attention in this direction has been devoted to neural connectivity information [18, 42–44], but databases with literature derived information from other modalities also exists, e.g. for electrophysiological features [45], and morphology and cytochemistry [46]. Most existing databases with literature-derived data contain qualitative or semi-quantitative information, and to date, no database of quantitative cellular parameters has been created. Such data are of broad interest to neuroscientists and a recent systematic review indicate that substantial variability exists in reported numbers [47].

The utility of public databases is evidenced by papers re-using and combining such data in new analyses [48–50]. For data shared through a repository to be as easy to use as possible, it is important that data are not only deposited, but integrated with other shared data. Avoiding the difficult challenges of combining disparate data, repositories sharing primary data [30, 35, 37] are based on highly standardized data acquisition pipelines to generate data. Repositories specialised to store and integrate primary data from *different* sources (e.g. NeuroMorpho [36], Cell-Centered database [41], ModelDB [34]) are relatively rare. Generalised

repositories like Zenodo (www.zenodo.org) and FigShare (www.figshare.com) allow scientists from any discipline to upload research data and receive a unique digital object identifier (DOI), but such data sets are not tagged with standardised and domain-specific metadata and may be virtually impossible to find without being directed to a specific link. This hampers discovery and re-use of data. In response to these challenges, Wilkinson and colleagues [51] introduced the FAIR principles for scientific data management and stewardship. These principles state that digital research objects (i.e. both data and the tools, algorithms, and workflows used to generate them) should be Findable, Accessible, Interoperable and Reusable (FAIR), both for humans and machines. The FAIR principles do not provide a standard or a specification of how data sharing should be implemented or which tools should be used, but are rather intended as high-level guiding principles. Specific standards tailored to different domains needs to be developed to promote FAIR sharing of data. Thus, there is a growing awareness that data can be considered a valuable research object when shared in accordance with the FAIR principles.

To be Findable and Accessible, data need to be assigned a unique and persistent identifier and to be catalogued in and possible to retrieve from a searchable resource. To promote Interoperability and Reusability, extensive annotation with metadata that in themselves follow FAIR principles is key, in addition to detailed provenance tracking and assignation of a clear usage licence. Enabling FAIR sharing of neuroscience data is an important objective of the European Human Brain Project (HBP), which aims to create an infrastructure for brain research that will advance neuroscience, medicine, and computing [13]. To this end, the HBP has created EBRAINS (www.ebrains.eu), a shared digital brain research infrastructure for the EU. Among other services, EBRAINS offers management, curation and sharing of neuroscience data with extensive metadata and unique DOIs assigned to every data set. EBRAINS metadata include basic metadata about contributors, licences and about the subjects used, location metadata about the investigated brain region(s), and metadata specific to the methods used. While the first category (basic metadata) provides the information needed to find and access the data, the latter two categories are critical to increase the interoperability and reusability of data. These types of metadata are essential to ensure that shared data can be integrated with other data in a meaningful way.

A challenge for any repository aiming to provide integrated data is that neuroscience data are heterogeneous, ranging both spatial and temporal scales [19]. Because information about anatomical position is arguably relevant to interpret neuroscience data across these modalities, such information provides an attractive basis for data integration efforts [52, 53]. Neuroscientists often employ reference

atlases in order to plan experiments and to determine the location of their data within the brain. Brain atlases are typically based on two- or three dimensional image data and show the names of and borders of various areas. Two types of metadata related to anatomical locations can generally be extracted from an atlas [54]: semantic metadata (region terms according to the terminology used in the particular atlas) and coordinate-based metadata (coordinates relative to internal or external landmarks). Such *anatomical metadata* extracted from atlases provide a natural piece of information by which different findings can be integrated. Integrating data through relating them to terms or coordinates in reference atlases is here referred to as *location integration*.

Reference atlases used in the field have evolved rapidly to provide an increasingly detailed reference. Traditionally, brain reference atlases used in murine research have been based on two-dimensional (2D) images of sections acquired from parts of or whole brains, with illustrations showing regional borders and names in relation to features in the underlying images. Coordinate systems that use skull-based landmarks as a reference are often applied to the illustrations, so that the atlases can be used to guide stereotaxic procedures. The rat brain atlases by Paxinos and Watson [55–59] and mouse brain atlases by Franklin and Paxinos [60–63] are the most detailed and widely used atlases available for murine brains. In addition, Swanson [64–67] has published detailed maps of the rat brain. Classical book atlases have long been invaluable tools in neuroscience laboratories. However, 2D atlases are limited by the distance between sections and distortions caused by processing and mounting of tissue. The utility for three-dimensional (3D) murine atlases in analysis of volumetric imaging data has been recognised for some time [35, 68–71], but they are also increasingly recognised as valuable tools for interpreting sectioned material [72, 73]. Indeed, sections that are not cut in one of the standard planes or with even slight deviations from these may not be comparable to the ones found in 2D reference atlases. The use of 3D brain reference atlases can mitigate such challenges [74]: in addition to providing superior resolution in the plane of sectioning compared to traditional atlases, 3D atlases can be re-sliced in any orientation. Essentially, this opens the possibility of transforming the atlas to fit data, instead of attempting to make the data fit the atlas, thus making 3D atlases ideal templates to which new data can be registered.

Creating a standardised 3D coordinate system to which new murine brain data can be registered was a central aim of the International Neuroinformatics Coordinating Facility's (INCF) digital atlasing programme [75, 76]. As a result, an atlas space defined by Cartesian coordinates with an origin related to internal landmarks, termed Waxholm Space (WHS), was conceptualised and implemented in

magnetic resonance imaging (MRI) data sets to create 3D atlases of the mouse [77] and rat [78] brain. The Waxholm Space atlas of the rat brain originally included 76 anatomical delineations across the brain [78], and has later been expanded to include detailed subdivisions of hippocampal [79] and auditory regions [80]. Volumetric atlases for the mouse have been available for some time [68, 69] but are becoming increasingly common and detailed with respect to the included delineations [81, 82]. The Allen Mouse Brain Common Coordinate Framework [32, 83] currently provides the most detailed volumetric atlas for the mouse brain. As openly accessible 3D atlases, the Waxholm Space rat brain atlas and Allen Mouse Brain Common Coordinate Framework are well suited to use for location integration [72] and in analytic pipelines [35, 84–86].

As elaborated above, several different atlases exist both for the rat and mouse brain. The 2D or 3D regional borders of an atlas are drawn on the basis of anatomical data (e.g. histological or immunohistochemical material [55–62, 64–67, 87], or magnetic resonance imaging [68, 69, 78–81]), and different features of brain anatomy may be visible depending on the data type. In addition, variable preferences exist for what to name regions and where to draw the boundaries between them. The many reference atlases that have been made for both the rat and the mouse brain thus vary in the terminology used and how regions are defined [88, 89]. Different views exist on how to divide even the most basic parts of the vertebrate central nervous system [27]. As subdivisions have become more fine-grained, confusion has increased. For example, the nomenclature for different regions of the murine prefrontal cortex has been subject to so much change and debate that comparing findings across studies becomes a serious concern [88, 90]. Similar discrepancies in definitions used to subdivide the human thalamus was demonstrated by Mai and colleagues [91], who asked several experts in the field to delineate the same section image. The discrepancies between regions as defined in different reference atlases have by some been termed the "brain atlas concordance problem" [39], and causes significant challenges for location integration.

Various efforts have been made to mitigate the brain atlas concordance problem. For example, Sugar and colleagues [44] created a "Rosetta table" of nomenclatures for the rat retrosplenial cortex, used to translate terms to a common nomenclature during collection of data for the temporal lobe database (www.temporal-lobe.com). Their approach involved careful consideration of the cyto- and chemoarchitectonic criteria described in the papers presenting each of the nomenclatures to create a table of synonym terms. Bota and colleagues [89] collated neuroanatomical nomenclatures and compared terms qualitatively in order to collect data for the Brain Architecture Management System (BAMS,

www.bams1.org). They assessed textual descriptions and graphical material provided with nomenclatures to determine qualitative spatial relationships between terms. However, these assessments were based on comparison of different 2D atlases where sections may be cut with different deviations from the standard planes [89]. Ideally, relationships between both terms and coordinates across different atlases would be defined based on one of the atlases being spatially registered to the other and taking deviations in cutting planes into account. This thesis is focused on murine brain research and using atlases to facilitate data integration in this field, but considerations made here should also be relevant for human atlases and registration of data to them, although the human brain shows much higher variability than murine ones.

To summarize, neuroscientists often struggle to find, interpret and combine the increasing amounts of data being produced in the field. There is also growing awareness that published information show substantial variability [47] and that current reporting practices do not support replication and reuse of data. New solutions for sharing, organising and annotating data are needed to integrate neuroscience data and make them Findable, Accessible, Interoperable and Reusable, in accordance with the FAIR principles [51]. Anatomical metadata from digital atlases has been emphasised as an important basis for data integration approaches in neuroscience [52, 53]. However, the challenge remains that there are several different views on how to define regions of the brain, and enforcing a single one in an evolving field might not be realistic nor desirable. Instead, there is a need for tools, standards, workflows, and sharing practices that support reuse and re-interpretation of data as new concepts and knowledge emerges. To meet this need, we asked the following questions:

- How can heterogeneous neuroscientific data be mapped to anatomical reference atlases, and what should be considered the minimum and best practices for reporting anatomical locations?

- How can quantitative cellular parameters reported in the literature be aggregated and integrated, and are reported numbers comparable?

- Can semi-automated methods efficiently deliver valid quantitative data that are comparable with previously reported data, and how can such data be shared to facilitate re-usability and transparency of analyses?

1.2 Aims

To address the questions listed above, one primary aim and three secondary objectives of this thesis were defined. **The primary aim was to establish new standards and approaches for mining, collecting, integrating, and sharing diverse neuroscience data from the literature, image repositories, and novel experimental work into common reference atlases of the rat and mouse brain.** To achieve the primary aim and to provide proof-of-principle for the novel standards and workflows developed in the project, we identified three secondary objectives:

- Establish standards and workflows for acquisition and reporting of semantic and spatial anatomical metadata
- Deliver a database of quantitative morphological parameters the basal ganglia of the normal adult rat and mouse brain based on information available in the literature and public repositories; and
- Perform a brain-wide analysis to describe numbers and densities of defined cell types across the rat and mouse brain, and establish how such data may be shared to facilitate their re-use

1.3 Structure of the thesis

The aims of this thesis are largely achieved through development of new standards and workflows. These are the primary results, which will be described in a summary of the results structured around the publications and data sets included in the thesis. The thesis does not include a "Methods" section, because the results section and procedures described in "Materials and Methods" of each paper together describe the methodologies in detail. Instead, the results section is followed by a section called "Methodological considerations", where I discuss considerations made during development of new standards and workflows, and how these differ from previous approaches. I will conclude with a general discussion focusing on the significance and utility of the results.

Chapter 2

Summary of results

The results described in this thesis are structured around three sub-projects. First, we established minimum and best practice recommendations for determining and documenting anatomical metadata. Secondly, a database of quantitative measurements from the basal ganglia of the rat and mouse brain was created, with data collected from the literature and from public repositories. Lastly, brain-wide data showing the distribution of two largely non-overlapping cell types in mouse brain were acquired, analysed, and shared in a way that enables integration with other data and re-use. The output of the different parts of the projects includes three papers and twenty-four data sets. In the following, I will briefly describe the papers and data sets delivered from each part of the project.

2.1 Navigating the murine brain

The first objective of the project was to establish standards and workflows for acquisition and reporting of semantic and spatial anatomical metadata. We asked how neuroscience data can be mapped to anatomical reference atlases, and what should be considered the minimum and best practice for reporting anatomical locations. We demonstrated that data obtained by several very different methodologies could be effectively registered to 3D atlases using new tools developed through the HBP. We defined two levels of recommendations for documenting anatomical locations: minimum practices which would allow for consistent interpretation and replication of studies, and best practices that would facilitate re-use and integration of data.

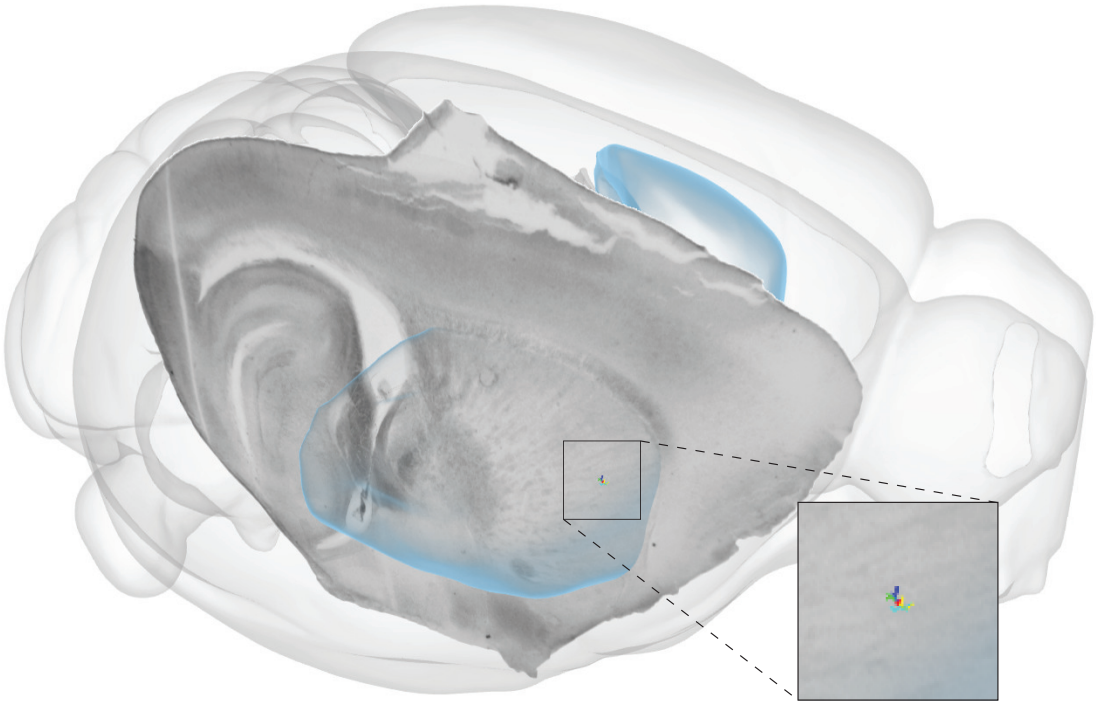


Figure 1. Navigating the murine brain. Figure showing the location of a reconstructed cholinergic neuron from the mouse striatum. The thick section (greyscale section image) was spatially registered to the Allen Mouse brain Common Coordinate Framework version 3. Based on identification of parts of the cells in this image, Neurolucida coordinates were translated to CCF coordinates and the reconstruction can be visualised and integrated with other data registered to the CCF.

2.1.1 Paper I: Best practice recommendations

To define new practices that can be fruitful to implement in the field, one must first acquire knowledge about the common practices. In our first paper, **"Navigating the Murine Brain: Toward Best Practices for Determining and Documenting Neuroanatomical Locations in Experimental Studies"**, we therefore asked how neuroanatomical locations are typically described and documented by researchers. The data from the literature survey presented in the first paper is provided as a supplementary file. Our results indicated that traditions vary considerably with the methodological paradigm. We observed that every third paper used an anatomical term without further definition, and that image documentation were seldom provided (only found in 34 % of papers). We then asked how this situation could be mitigated with relatively simple measures, and proposed a set of minimum requirements for reproducible reports of anatomical locations. We intended these requirements to increase consistency and reproducibility of location reports. We further demonstrated additional steps that can be taken in order to more completely document anatomical locations in a way that facilitates comparison and integration with other data.

The example data sets described in the paper are not outputs of the current project, but were rather used here to exemplify the workflows for spatial registration. Most of these are available as shared data sets:

- The neuronal reconstruction of a striatal interneuron is available through a data set hosted at EBRAINS [92]
- Parvalbumin-stained horizontal section images are available through a data set hosted at EBRAINS [93]
- Parvalbumin-stained sagittal section images are available from the Allen Institute (http://mouse.brain-map.org/experiment/show/75457579) [31]

The remaining data sets used to exemplify the workflows (i.e. the electron microscopy [94] and electrophysiology [73] data sets) were kindly shared with us by colleagues.

2.2 Integrating published quantitative cellular measurements

The second objective of the project was to deliver a database of quantitative morphological parameters from the normal murine basal ganglia. We asked how quantitative cellular parameters reported in the literature can be aggregated and combined, and whether reported numbers can be considered comparable. To address this question, we designed a database and established a new workflow for semantically registering published information to brain atlas regions. We found and organised more than 1000 published quantitative estimates from the basal ganglia in this database, which we delivered as a openly accessible data set. Lastly, we used the database to show high variability in reported estimates, common lack of metadata, and that few replications are available for most cell types.

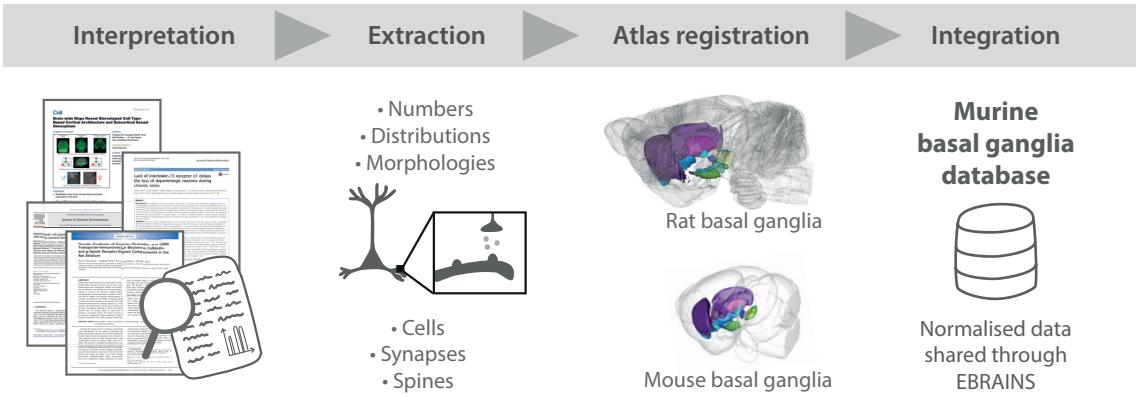


Figure 2. Integrating published quantitative cellular measurements. Figure illustrating the creation of the Murine basal ganglia database. Published information was identified and carefully extracted. Parameters of interest included numbers and distributions of cells, synapses and spines, as well as cell morphologies from rat and mouse basal ganglia regions. These data were semantically registered to common reference atlases and integrated in a relational database called the Murine basal ganglia database.

2.2.1 Paper II: Database of basal ganglia cellular measurements

We used the guidelines for minimum and extended documentation of anatomical location [72] as a starting point to collect anatomical metadata about regions of interest described in the literature to be included in our database. However, simply knowing how a location has been documented was not enough. The comparability of that location to one reported in another paper is also essential to compare results. A major challenge was the variable definitions of similar region terms, as different

researchers prefer different reference atlases or even use custom definitions. We devised and implemented a step-wise workflow to compare regions of interest in different reference atlases, which formed the basis semantic registration to atlas through translation from one terminology to the other. Specifically, we translated all region terms found in the literature to corresponding ones in the murine atlases used in EBRAINS. We also collected metadata about experimental animals, methodological parameters and analytic procedures. We present our database (called the Murine basal ganglia database), in the second paper of this thesis, **"Database of literature derived cellular measurements from the murine basal ganglia"**. The paper is a Data Descriptor, which describes the creation and content of the data set, provides examples of how the content may be used, and gives perspectives on the comparability of numbers in the database. Importantly, we describe how the database may be used to update content in the future, facilitating long-term relevance of the data collection.

2.2.2 Data set: Database of basal ganglia cellular measurements

The database content can be accessed through a data set in the EBRAINS Knowledge Graph, under a CC-BY licence [95]. This data set consists of three parts: 1) the Murine basal ganglia database (shared both in the original Microsoft Access .accdb format, and as .csv files for each table); 2) an empty version of the MS Access database, with explanatory input forms through which new data can be entered; and 3) an Excel workbook that may be used in a simplified workflow to collect more data. The latter two parts of the data set are intended to facilitate collection of more data in the future.

2.2.3 Data sets: Spatial co-registration of murine atlases

To relate all the data collected for our database to the rat and mouse brain reference atlases used in EBRAINS [78–80, 83], it was first necessary to know how regions defined in other reference atlases relate to the ones seen in EBRAINS atlases. Using the QuickNII tool (RRID:SCR_016854) [73], we spatially registered the most commonly used rat and mouse brain atlases in the field [55–57, 60–67, 96] to the MRI/DTI template of the Waxholm space atlas of the Sprague-Dawley rat brain (WHS) (version 1.01, RRID:SCR_017124,) [78] or the serial two-photon tomography (STPT) template of the Allen mouse brain Common Coordinate Framework (CCF, version 3) [32, 83]. This allowed us to directly compare regions in the WHS or CCF

with those in other atlases. We made qualitative and semi-quantitative descriptions of relationships between basal ganglia regions in different atlases, which are shared through separate data sets for rat [97] and mouse [98]. Beyond these descriptions, the spatial co-registration information might be useful for other projects where differences between atlases needs to be considered. We have therefore shared the spatial metadata defining these co-registrations as .xml files generated in QuickNII. Fourteen data sets are available from the EBRAINS Knowledge Graph, each one containing the spatial metadata defining the co-registration of one atlas to the WHS or CCF template [99–110].

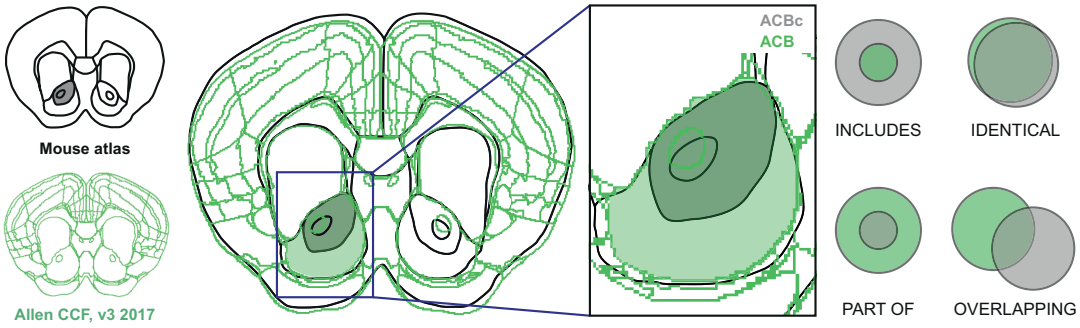


Figure 3. Co-registration of reference atlases. Plates from the most commonly used reference atlases were registered to EBRAINS reference atlases. This allowed us to determine whether regions in other atlases were identical to, included, were part of, or overlapped with regions in EBRAINS atlases.

2.3 Quantifying the murine brain

The third objective of the project was to perform brain-wide analysis to describe numbers and densities of defined cell types across the rat and mouse brain, and establish how these data may be shared to facilitate their re-use. We asked whether semi-automated methods can efficiently deliver valid quantitative data, whether these data are comparable to published information, and how they may be shared to facilitate transparency and re-use. We performed brain-wide quantitative analysis of parvalbumin and calbindin cells using the recently published QUINT workflow [86], and showed trends comparable to literature data. Our open sharing of all generations of these data sets, as well as metadata needed to interpret and replicate analyses, facilitate re-use and thus long-term relevance of the data.

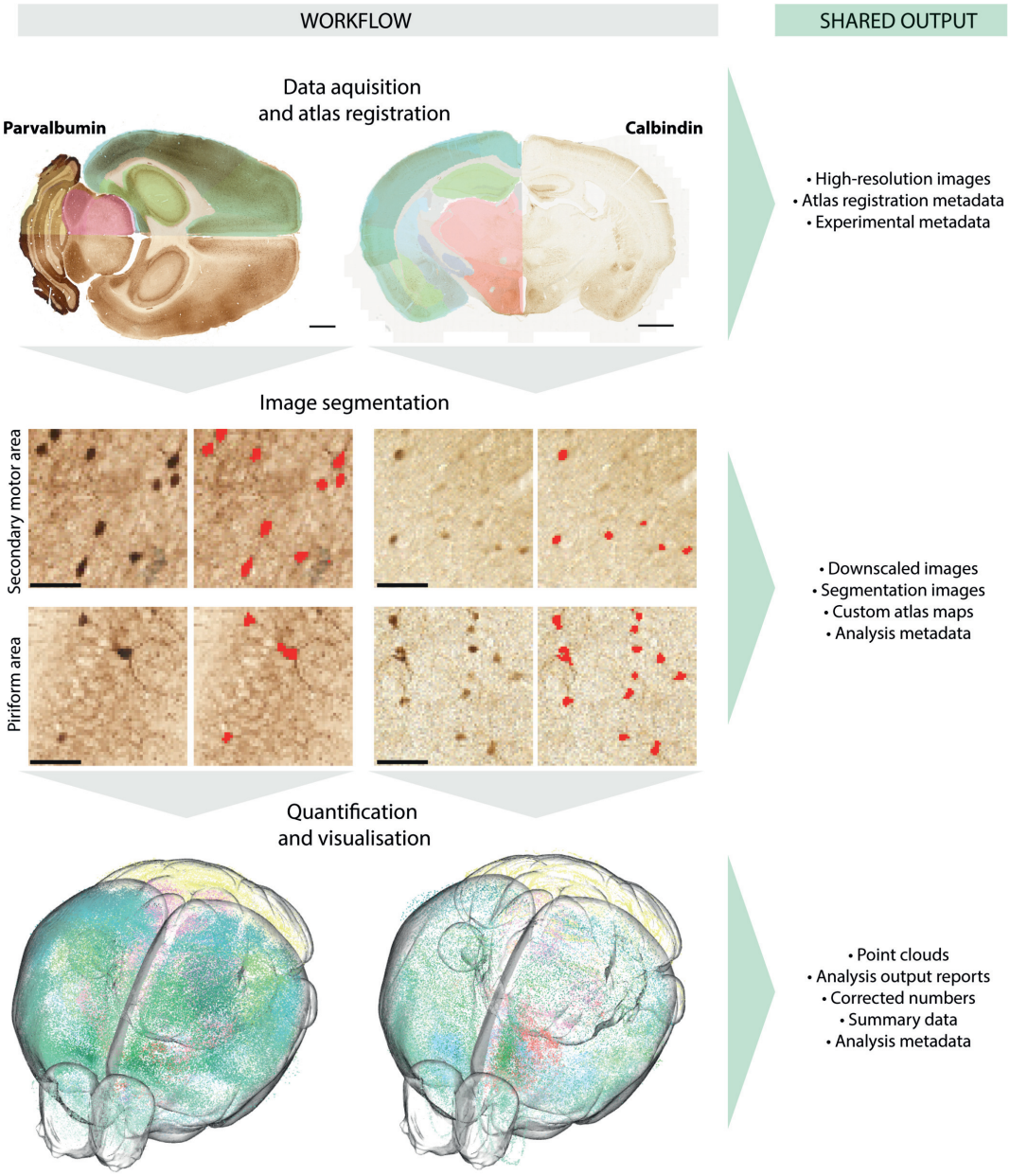


Figure 4. Quantifying the murine brain. Figure showing the workflow used to generate brain-wide quantitative data for parvalbumin and calbindin cells (left panel) and the output shared from each part of the workflow to facilitate transparency of analysis and re-usability of results.

2.3.1 Paper III: Brain-wide analysis of parvalbumin and calbindin cells

In the third paper, we present the results from a brain-wide analysis of data sets immunohistochemically stained for parvalbumin and calbindin-D28k (here referred to as calbindin). First, we mapped and compare the numbers of calbindin and parvalbumin cells in the mouse brain, showing that these cell types distribute according to largely complementary patterns. Parvalbumin neurons were dominant in regions related to sensorimotor functions and navigation, while calbindin neurons were more abundant in areas associated with modulation of behavioral state. Secondly, we analysed the distribution of parvalbumin cells in the rat brain as well, and compared the number, density and distribution of these cells within and across hippocampal regions in the two species. We compared our findings to a previous brain-wide study of parvalbumin cells in the mouse brain [85] and to several other publications. This comparison showed similar trends across regions in different data sets, although the numbers varied. All the primary and derived data collected during this analysis were openly shared, with accompanying data descriptors detailing how they may be re-used in new analyses. In addition to providing the first brain-wide quantitative analysis of immunohistochemically stained parvalbumin and calbindin neurons, the third paper thus provides a demonstration of how data may be analysed and shared to enable interoperability and re-usability of the data.

2.3.2 Data sets: Brain-wide distribution of parvalbumin cells

Brain-wide series of parvalbumin stained sections from the rat and mouse were acquired and analysed. The raw data includes high-resolution images of the sections, with one data set containing the rat brain data [111] and one containing the mouse brain data [112]. The images were spatially registered to the Waxholm Space atlas of the Sprague-Dawley rat brain and the Allen Mouse brain CCF (version 3, 2017 edition), respectively, and the spatial metadata describing these registrations are shared together with the raw data. We used the data in a brain-wide analysis of parvalbumin neuron quantities. The derived data are shared through individual data sets for the extracted rat [113] and mouse [114] brain data. These data sets

include downscaled images used for segmentation of cell bodies, segmentation images and custom-cut atlas maps for each section, and quantitative results obtained in our analysis.

2.3.3 Data sets: Brain-wide distribution of calbindin cells

Calbindin stained sections from the whole mouse brain were generated and analysed. As for the parvalbumin data, the raw data consists of high-resolution section images [115] shared together with the spatial metadata describing their registration to the Allen Mouse brain CCF. Derived data generated through our analysis, including the same components as for the extracted parvalbumin cells, are also shared through a separate data set [116].

Chapter 3

Methodological considerations

The results described in this thesis include new standards and workflows, as well as new applications of existing methods. Thus, methodological choices and considerations taken during the course of the work have been key to achieve the results, and are discussed in the following. I will first present and discuss the evaluation criteria for anatomical metadata described in the first part of Paper I, which were the basis for our minimum and best practice recommendations. Secondly, I elaborate on the spatial and semantic atlas registration procedures, which are important throughout all three papers in the thesis; I also explain aspects of the atlas co-registration used as a basis for semantic registration by translation of terms in Paper II. I go on to elaborate on the decisions made while designing the Murine Basal Ganglia Database (Paper II) and compare this design to another public database. Finally, I discuss some important aspects of established counting methods and how these compare to the solutions for cell segmentation and quantification used in our Paper III.

3.1 Evaluation of anatomical metadata

In Paper I, we systematically assessed documentation provided in the literature and defined a set of standard reporting practices for anatomical metadata. To do this, it was first necessary to define the different ways that regions of interest (ROIs) may be documented. By exploring the literature and considering approaches used in the HBP, we established an evaluation system consisting of nine factors, summarised in Box 1.

Box 1 | Factors of the documentation score

Description of region of interest: How the authors describe their definition of the region of interest. One or more of the following options:

1. ROI is mentioned in text without further explanation
2. ROI is defined by distance to landmark
3. ROI is defined by a reference atlas
4. ROI is illustrated by use of image from a reference atlas
5. ROI is illustrated by image or principal diagram
6. ROI is described by cyto- or chemoarchitectonic criteria, or electro-physiological features

Sectioning: Whether or not sections (≤120 µm) or slices (>120 µm) have been cut

Staining: Whether or not staining of sections or slices have been performed

Coordinates: Whether or not coordinates related to the ROI is given. One of the following options:

1. Coordinates of sections or slices
2. Point coordinates of ROI
3. Point coordinates within a larger ROI

Image documentation of own material: Whether or not images of the analyzed material is included. One or more of the following options*:

1. Image(s) of dissection procedure
2. Section images showing ROI only
3. Section image showing part of section
4. Section image showing half a section
5. Section image showing whole section

Anatomical annotations: Whether or not annotations of regions are superimposed on image(s). One of the following options:

1. Annotations without anatomical borders
2. Delineations of (sub)regional boundaries from anatomical features
3. Delineations of (sub)regional boundaries from histoarchitectonic features

Number of sections: How many section images that are provided. Options are grouped as 2-5, 6-10, 11-20, and 21+ sections.

Spatial registration to atlas: Whether or not material has been registered to a reference atlas, and the image material with corresponding atlas information is made available. One of the following options:

1. Superimposed standard atlas diagram (2D)
2. Superimposed diagram from 3D atlas

Transformation: Whether or not the spatial registration to atlas has been optimized to fit the section images. One of the following options:

1. Linear (rigid and/or affine) transformation
2. Non-linear transformation

Box 1. Factors of the documentation score. Nine factors related to documentation of anatomical locations were recognised and are described here (see text for details).

The first factor is description of the ROI. This concerns how and to which degree the authors describe their definition of the ROI. We considered six types of descriptions (Box 1). For a ROI to be considered defined by use of a reference atlas (description type number four), we required that a clear statement was made that an atlas had been used for defining the ROI, including a proper reference.

The second factor is sectioning, and concerns whether or not sections or slices have been cut as part of the experimental procedure. Sections or slices can be compared to illustrations in reference atlases and thus makes recognizing ROIs convenient. The third factor concerns histological and immunohistological staining, which greatly aids in recognizing ROIs, and this was only assessed for experiments where sections or slices have been cut.

The fourth factor is coordinate based information, of which we distinguished three types (Box 1).

The fifth factor is image documentation. In this context, only images that aid interpretation of anatomical location was considered image documentation. Information on the coverage of image documentation was considered of interest. The influence of image coverage on interpretation of anatomical positions is shown in Figure 2 of Paper I, where two researcher's interpretation of a position gradually converges on a consensus as the coverage increases. We distinguished four types of image documentation (Box 1). The sixth factor is anatomical annotations, and applied only to those reports that provided images. We required annotations to be superimposed on images to be considered anatomical annotations (section illustrations on the side of an image were considered a description based on an illustration) and that it was clear which annotation represented the region of interest. This typically meant that the region name should be superimposed on the image, but other indications (e.g. color coding) were also sporadically observed. We distinguished three types of anatomical annotations (Box 1). The seventh factor is the number of section images provided, and here we only considered sections that were documented in the paper or otherwise made available to the reader.

Spatial registration to atlas is the eight factor, and concerns whether the material has been registered to a reference atlas. This is a key strategy to integrate data in the HBP, but was rarely adopted in the literature surveyed here. We distinguished between registration to two- or three dimensional reference atlases. Various transformations may be performed to make the atlas fit the material (or vice versa). Thus, the ninth and last factor is transformation. Such transformations

may be linear (including rigid and/or affine transformations) or non-linear.

3.2 Integration of data through atlas registration

Our survey of how neuroanatomical locations are determined and documented in the literature revealed that basic descriptions are very often lacking and indicated large variations in approaches across methods. This creates challenges for interpretation, comparison, and integration of findings across studies. In the work presented in this thesis, we therefore establish new worfklows for registration of data to atlas and for comparing data related to different atlases, facilitating location integration of data. In the following, I first consider some important aspects of atlas registration procedures, and secondly elaborate on methodological aspects of our co-registration and comparison of different atlases.

3.2.1 Spatial atlas registration

Spatial atlas registration is the process of defining anatomical location of data in an atlas by use of the atlas coordinates. We used the QuickNII software [73] as the basis for the spatial registration procedures described in this thesis, which requires that some form of image material, preferably section images, are available. QuickNII allows the user to cut three-dimensional atlases in any plane of orientation to match the cutting angles of a given section. In our experience, almost all sections have a slight deviation from the standard plane that is easily detected when fitting the atlas to the section. Such deviations may have considerable effects on the interpretation of anatomical location [72]. Other approaches to account for deviations from the standard anatomical planes have been based on manual interpretation of multiple reference atlas plates [48], a time consuming and error prone process. Generation of custom-cut atlas plates and linear stretching of the atlas allows convenient assignation of spatial and semantic atlas information to new data. Few if any other tools with a graphical user interface offer such features [73]. Thus, if section images are available, these can be directly registered to the atlas for the relevant species using QuickNII. This is particularly relevant for methods that generate such images, e.g. tract tracing, immunohistochemistry, and *in situ* hybridisation, but we also show that this approach can be a realistic alternative for methods where such sections do not constitute the primary data (e.g. electron microscopy, neuronal reconstruction) [72].

The QuickNII tool does not allow non-linear adjustments to be made to the atlas plates. Distortions to sections as a consequence of cutting, staining,

and mounting processes are common [74]. Determining the exact cutting plane and applying linear transformations using QuickNII is sufficient for qualitative or semi-quantitative analyses [117] or for determining the position of one or a few sampling sites [94]. However, for brain-wide quantitative analyses, more accurate definitions of regions across the brain is essential. Registration accuracy can be significantly improved by non-linear adjustments, and further refinement of the linear registration from QuickNII can be made using the recently developed VisuAlign tool (RRID:SCR_017978; Puchades M, Csucs G, Yates S, Leergaard TB Bjaalie JG, personal communication). In Paper III, we demonstrate the utility of QuickNII and VisuAlign to create customised and non-linearly refined atlas plates to be used in semi-automatic, brain-wide analyses. Figure 4 illustrates the workflow for spatial registration using QuickNII and VisuAlign.

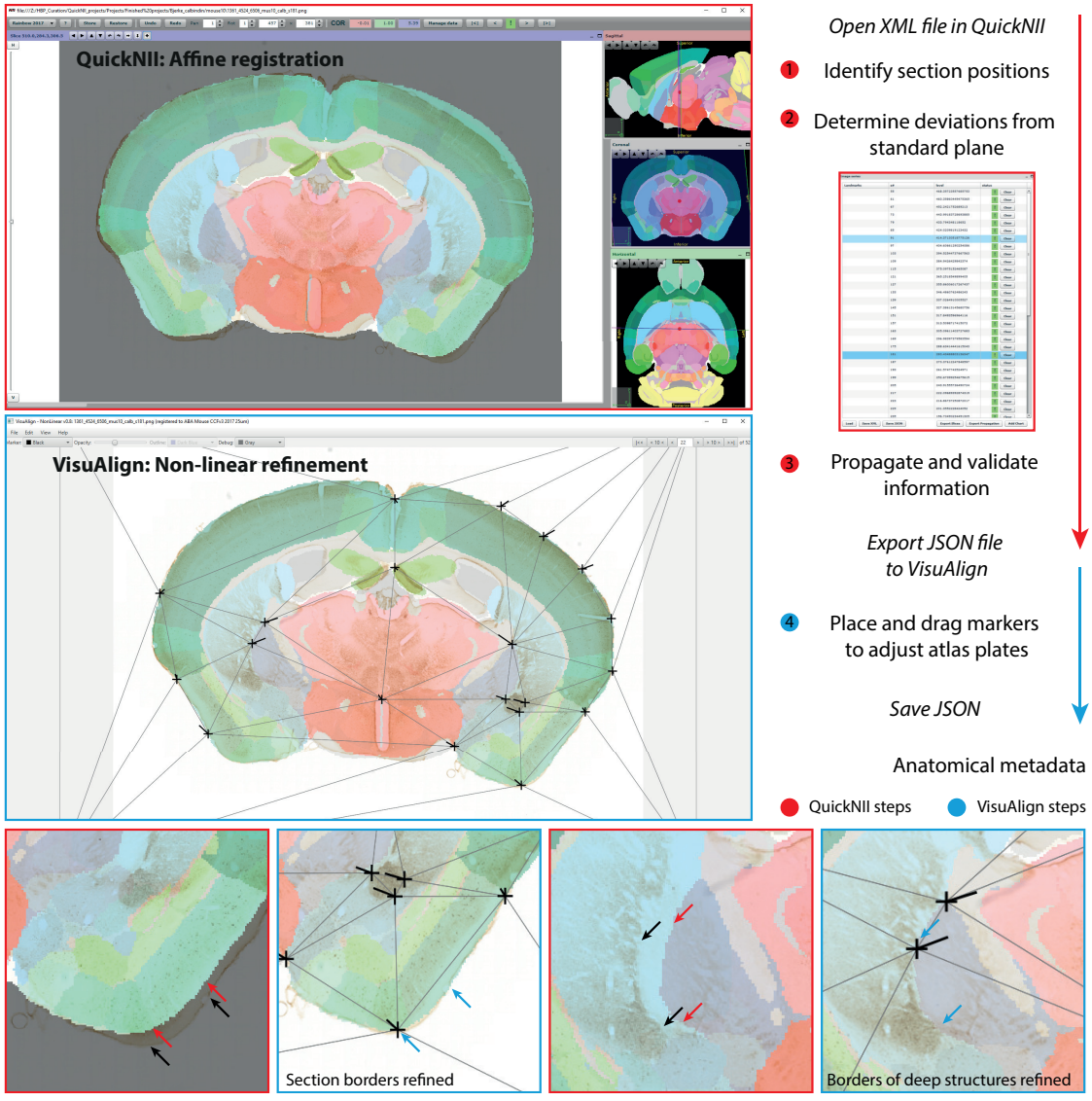


Figure 4. Spatial registration with QuickNII and VisuAlign. A global registration in QuickNII is achieved by identifying the position of each section and

deviations from the standard plane. Linear transformations and rotations can be made to the atlas maps. This is followed by non-linear refinement of each atlas plate in VisuAlign. The upper panel shows the steps taken in each software to achieve the registration. In the lower panel, magnified views of the registration result after QuickNII and VisuAlign registration steps are shown for the cortical surface (left) and caudoputamen-globus pallidus border (right).

3.2.2 Semantic atlas registration

Generation of sections or slices is not an integral part of all methods used in neuroscience. When image material is not available, only anatomical terms can be used in order to connect data to the relevant locations in atlas. This process of defining anatomical location of data in an atlas by use of the atlas terminology is here referred to as semantic atlas registration.

Semantic registration by translation of terms

A key prerequisite for integrating data in the murine basal ganglia database was to establish a workflow for translating the myriad of names, abbreviations, and lexical variants used in the literature to consistent terms used in EBRAINS atlases. When describing anatomical regions in articles, researchers might refer to terms found in a reference atlas. Alternatively, and quite commonly, terms might be used that are not consistent with, or without specification of, an atlas. These are here referred to as custom terms. For atlas terms, we based our translation on how those terms spatially correspond to those in EBRAINS atlases (see below for details). For custom terms, translation was based on careful consideration of the documentation presented in each paper.

It is well known that the meaning of an anatomical term often varies depending on the consulted atlas, book, or researcher (for a summary of various views on neuroanatomy, see [27]). Efforts have been made previously to compare terms and definitions from different nomenclatures, largely based on linking synonym terms [44]. For the basal ganglia regions, our atlas comparison analysis clearly shows that the "identical" relationship (which would indicate that two terms are synonymous), would not suffice to describe the relationships between atlas terms. More complex relationships between atlas regions were defined by Bota and colleagues [89], although based on comparison of 2D material. The authors did note that specification of relationships ideally "could be accomplished rigorously in a high-resolution 3D, reslicable computer graphics model for a particular species" ([89], page 1). Our approach, based on co-registration of atlases using the QuickNII

tool, indeed allows one atlas being custom sliced to fit another. This allowed us to make more direct comparisons of delineated structures than has been possible for murine brain atlas region previously.

Considerations and metadata related to semantic registration

There are several important aspects related to the decisions made during semantic registrations. These are summarised in Figure 5 and discussed in the following.

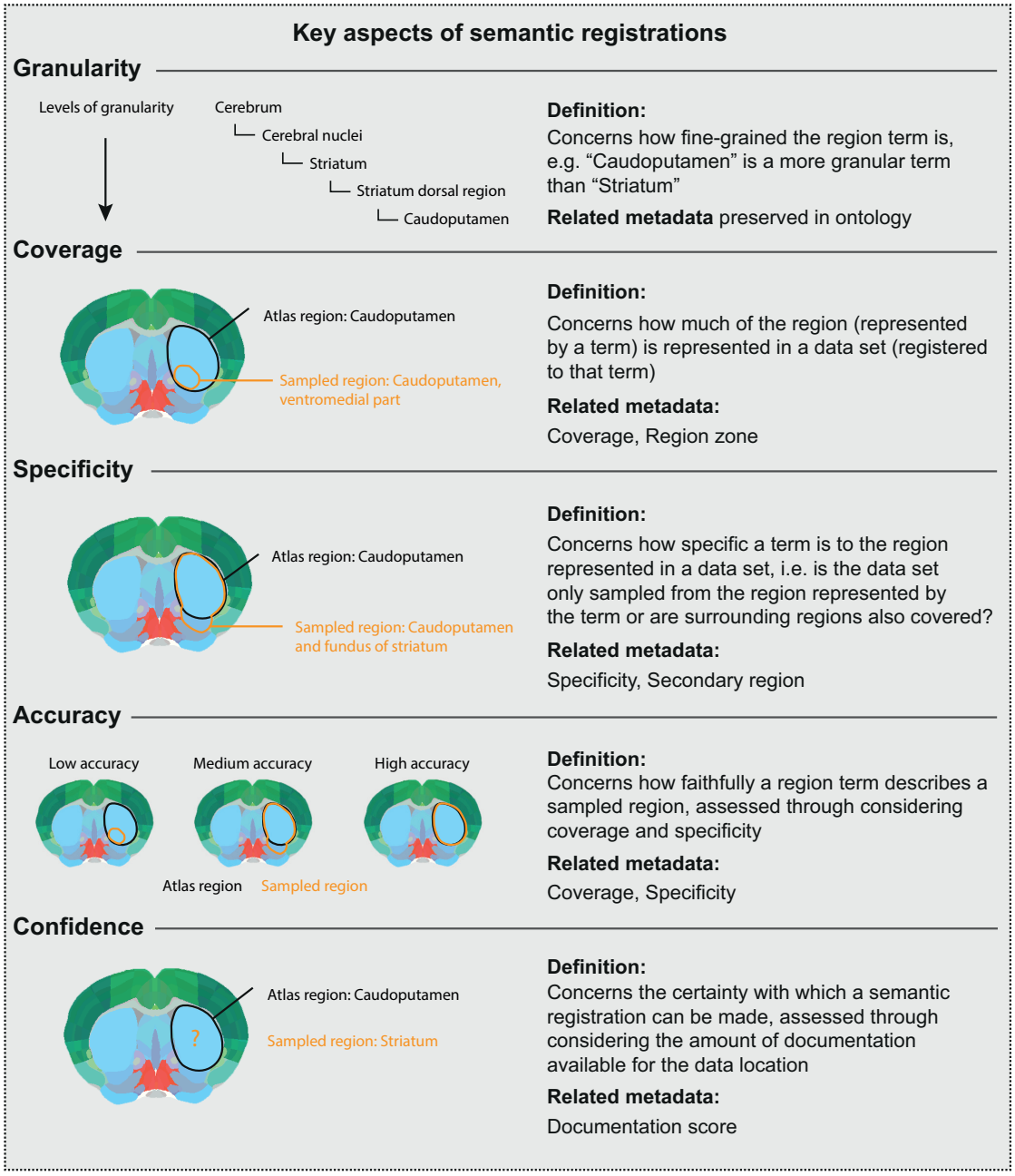


Figure 5. Key aspects of semantic registrations. Throughout the figure, black text and illustrated areas represent an atlas region to which a data set

(orange text and illustrated areas) has been semantically registered. Concepts of granularity, coverage, specificity, accuracy and confidence are defined and illustrated. Metadata for each concept, as stored in the Murine Basal Ganglia Database, are listed. Information of the original sampled region is also stored as metadata in tables connected to the semantically registered region.

The accuracy of a semantic registration is determined by how faithfully the atlas term represents the region from which the data were obtained. The data may not fully cover the region represented by the atlas term (e.g. data may originate from only part of the region termed caudoputamen in an atlas). Conversely, the atlas term may not be specific to the data (e.g. data may primarily originate from the region termed caudoputamen in the atlas, but with some parts extending into the nucleus accumbens). Together, the coverage and specificity indicate the registration accuracy. A semantic registration will also have a certain level of confidence (depending on the amount of information available on which to base the registration). When the coverage and specificity is unknown, the confidence is low (the registration may be accurate, but it is not possible to know).

Is full coverage or high specificity more important during semantic registration? Or in other words, how fine-grained ("granular") should the region term be? The appropriate granularity of the terms used for semantic registration needs to be determined based on the accuracy and confidence of the resulting registration. A term should not be so granular that specificity is low (i.e. the data covers the entire region but might originate from surrounding regions as well). At the same time, the term should not be so coarse that the data have very low coverage (i.e. the data originate from a very small part of the region only). In many cases, reducing coverage is preferable to reducing specificity, thus ensuring specificity rather than using the most granular term. For example, a data set of reconstructed cells obtained primarily from the caudoputamen but possibly including cells from the nucleus accumbens could still be correctly registered to the striatum. However, for cell counting data, and especially for total number estimates, both specificity and coverage are critical factors affecting comparability to other estimates. In these cases, we opted for a balance where a less granular term was only selected if this did not significantly affect the coverage.

Metadata in the Murine basal ganglia database describing the semantic registration include coverage (whole region, part of region, or unknown) and specificity (specific, non-specific, or unknown). For cases with "part of region" coverage, we added information about the covered parts in a field called "Region zone" (e.g. "dorsolateral part"). For registrations with specificity set to "non-specific", a field called "Secondary region" was included to indicate other regions that were covered by the data. Any further comments about the coverage

and specificity of a registration was added to a "Collectors comment" free-text field. Information about the documentation found for each region of interest was also stored, in accordance with the documentation factors defined and listed above (see "Creating standard reporting practices for anatomical metadata"). The documentation information was used to generate a "Documentation score" ranging from 1-10. This score will not tell how accurate the registration is, but simply how much information was available about a region of interest. Thus, the documentation score will give an impression of the registration confidence. For example, for the registrations with high documentation scores, the coverage and specificity could typically be determined.

3.2.3 Atlas co-registration and comparison

As described in the introduction, there are several atlases available both for the rat and mouse brain that vary in how they name and define regions of interest. In order to translate atlas terms from one terminology to another, we spatially co-registered several of the most commonly used murine brain atlases to the rat and mouse brain atlases used in EBRAINS using QuickNII. For each basal ganglia region in different reference atlases, we qualitatively and semi-quantitatively evaluated the comparability to region(s) in EBRAINS atlases. This information was used when translating terms during semantic registration, as described above. Here, I will consider important methodological aspects of our comparability evaluation, and discuss how our approach compares to other solutions to the brain atlas concordance problem.

Evaluation of differences observed between atlases

Our approach involved making custom cuts of volumetric atlases to fit two-dimensional ones. This allows direct comparison of delineations and minimises the effect of slight deviations from standard cutting planes that might be seen in section images used in 2D atlases. When observing differences between the atlas plates, several possible reasons needed to be considered:

- **Inaccuracy in the registration.** Since our approach involved linear deformations only (see section on "Spatial registration to atlas" above), there were necessarily some inaccuracies in the registration. When comparing regions in various atlases using QuickNII, inaccuracies were mitigated by adjusting the registration to give a better local fit for each investigated region. This was done manually, and did not pose a significant challenge to the

procedure employed here. Nevertheless, limited accuracy of the registration constituted one possible reason for observed differences.

- **Animal differences.** Observed differences between atlases may reflect differences in the anatomy of the brains used for different atlases: while adult male brains are traditionally used [32, 57, 66, 78], animals of different strain or age might be used in different atlases. The size of different regions might vary differently across ages and strains [118, 119].

- **Delineation differences.** Lastly, differences observed between atlas plates could indicate differences in the delineation criteria. These are the differences we were interested in capturing in our comparison. A combined assessment of the spatially co-registered atlases, the terminology, and the criteria defined for each atlas region was necessary to determine whether an observed difference was likely to be due to differences in delineations.

Rat brain atlases				
Atlas	**Sex**	**Strain**	**Age**	**Weight**
WHS	Male	Sprague Dawley	80 days	397.6 g
Paxinos stereotaxic	Male	Wistar	N/A	270-310 g
Paxinos MRI	Male	Wistar	N/A	300 g
Swanson	Male	Sprague Dawley	80 days	315 g
Mouse brain atlases				
Atlas	**Sex**	**Strain**	**Age**	**Weight**
AMBA CCF v3	Male	C57BL6*	56 days	N/A
Franklin and Paxinos	Male	C57BL6	90 days	26-30 g

Table 1. Subjects used for the atlases assessed in our comparability study. Atlases: WHS, Waxholm space atlas of the rat brain; Paxinos stereotaxic; stereotaxic rat brain atlases by Paxinos and Watson; Swanson; stereotaxic rat brain atlases by Swanson; AMBA CCF, Allen Mouse brain Common Coordinate Framework (version 3); Franklin and Paxinos, stereotaxic mouse brain atlases by Franklin and Paxinos. *AMBA CCF v3 is based on a population averaged template derived from serial two-photon tomography (STPT) of 1675 mice [83].

Comparison of atlases was performed using semi-quantitative scoring of region comparability (described in Paper II). The results of this approach include descriptions of relationships based on neuroanatomical knowledge [97, 98], as well as identification of areas where opinions in the field are variable. Quantitative comparison was not attempted due to the registration inaccuracies discussed above. At the time of our atlas co-registration, the VisuAlign tool was not yet available. A future improvement to these data sets might therefore involve a nonlinear refinement

of the registration, in which case a quantitative approach to atlas comparison might be more feasible. This could also be used as a basis for translation of coordinates across atlases. Our co-registration of multiple atlases to a common, custom-sliced 3D template, will allow such translations to be more efficient and less dependent on manually consulting multiple plates from each atlas than has been possible previously [120].

3.3 Building a database of quantitative cellular measurements

The second objective of the project was to deliver a database of quantitative morphological parameters reported in the literature. Derived data found in the literature presents some particular challenges for integration: they may be presented across text, tables and figures, and described by use of different concepts and units of measurements. Nevertheless, the body of literature constitutes a large and important source of collective knowledge. Strategies for integration and comparison of individual pieces are therefore crucial and will likely be useful also to integrate data from novel experimental work.

3.3.1 Design strategy

Both bottom-up and top-down strategies were required to design the database to be used. From a bottom-up perspective, it was important to consider the parameters that are commonly reported and how they are normally presented, as well as the methods are typically used and how these may be described. Continuous exploration and testing with actual use-cases was needed in order to refine the design, to ensure that the system was compatible with the available literature data and metadata. However, the actual metadata found in the literature is not standardised and may be incomplete. From a top-down view, it was important to consider the information that was desired in order to represent data, and the ways in which data and metadata could be structured and presented in a unified way. By combining these approaches, it was possible to establish strategies to make use of what is actually present in a more structured way. Figure 6 shows a simplified Entity-Relationship Diagram of the database, and the most important tables and metadata stored in them are discussed in the following.

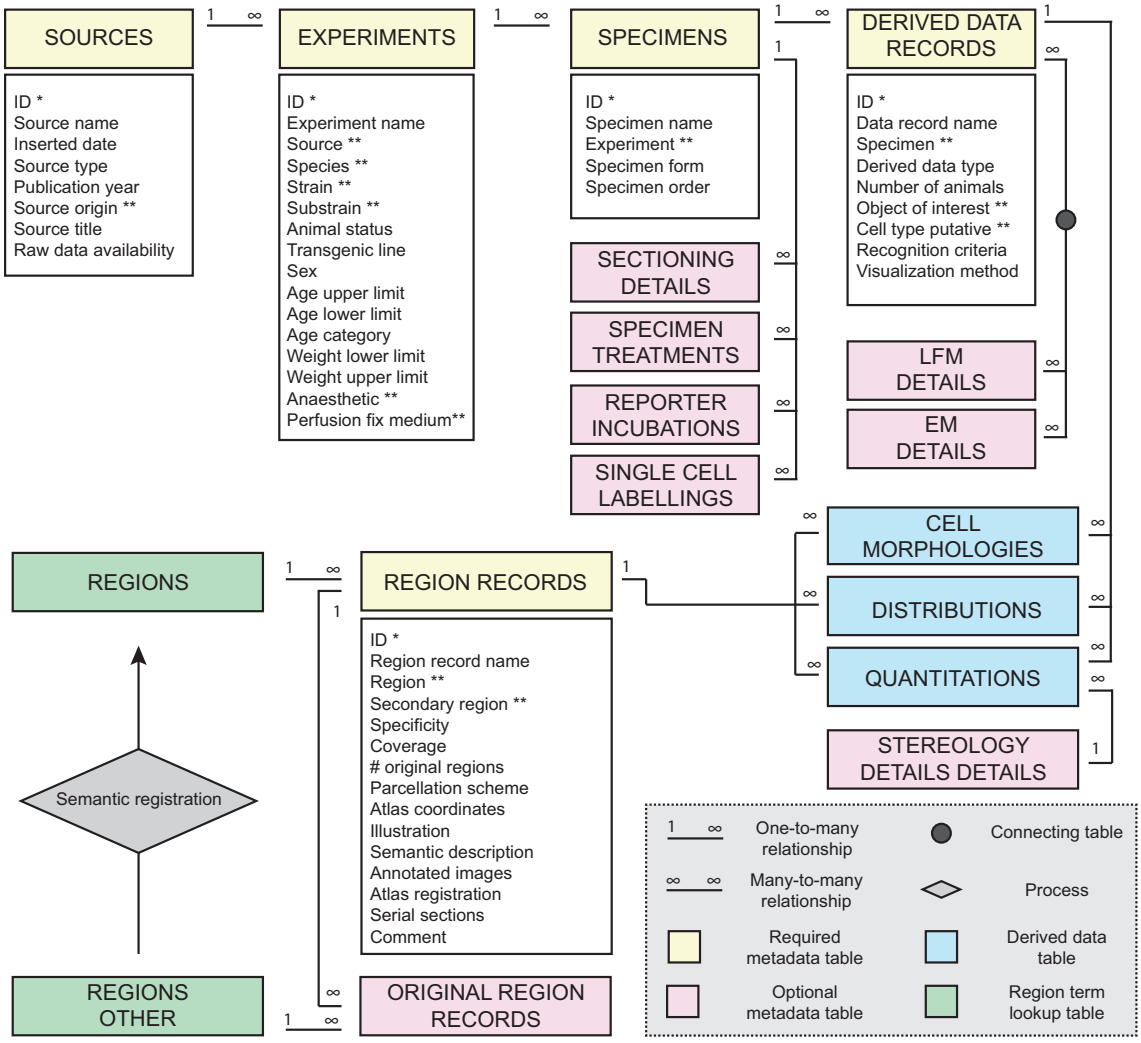


Figure 6. Simplified Entity-Relationship diagram for the Murine Basal Ganglia Database. Tables in the database are illustrated. Main metadata tables containing information about the source, experiment, specimens, derived data records and region records are shown in yellow; these are required tables to which data is always entered. Detailed methodological metadata are filled into the tables illustrated in pink, depending on the methodology used. The derived data are entered to the tables illustrated in blue. Metadata about the original region term is stored and translation to standard terms is made. Lookup tables with information about standard regions ("Regions") and the terms found in the literature ("Regions other" are illustrated in green.

Database structure and included metadata

We structured our database following the logic of how experiments are typically performed and how they are typically reported. A source (literature or repository) can include information from multiple experiments. Since derived data in the literature are by far most commonly given as group estimates, it is generally not possible to map specific results to any ID for the animal that it came from. Therefore,

our database was not designed to represent the individual animals used in a source, but rather the group of animals used in an experiment. An experiment, in this context, was thus defined by the group of animals it was performed on.

Each experiment could give rise to multiple specimens. A specimen was in this context defined as a group of tissue samples with similar form and which has undergone similar treatment. For example, a group of brains used in an experiment would constitute one specimen. If tissue sections were cut from these brains, they would together constitute another (secondary) specimen. These specimens were not linked directly, but would both be linked to the same experiment and annotated as "primary" and "secondary" in another field. There could also be "tertiary" and "quarternary" specimen, e.g. pieces dissected from a section or ultrathin sections for electron microscopy.

During an experiment, various processing steps are usually applied on the specimen to be analysed. Some treatments (e.g. fixation) are typically done while the brain is still intact, while others (e.g. staining, permeabilisation) are performed on sections of the tissue. One strategy for representing tissue processing metadata in the database would be to attempt capturing the "typical" experiment, with separate tables for treatments of the various forms of tissue (brains, sections, ultrathin sections) and fields for information about each step in the protocol. While neuroanatomical experiments to a large degree follow standard methods, there are variations that quickly made such an approach problematic. For example, embedding brains in sucrose is a typical step when cryosectioning is to be performed, but not when a vibratome is to be used for cutting. Fixation steps may be performed on the intact brain, or on sections, or both. To achieve the necessary level of flexibility, separate specimen records were kept for the different tissue types. Information about the treatment that was used on a specimen is stored in a separate, connected table, with fields for the chemical solution used, as well as the time, temperature and purpose of the treatment. For the chemical solutions, we did not attempt to capture each component of every solution separately, or information about concentrations. While this would be possible, it would involve creating several more tables and would require significant extra time during data entry. Instead, we simply included a lookup table including all the chemical solutions, with each one named by its components separated by dashes (e.g. paraformaldehyde-glutaraldehyde). Note, however, that treatments with antibodies and RNA probes were stored in separate tables where information about concentration can be added.

A specimen can give rise to one or more derived data record. We defined a derived data record a set of analytic results derived from the same specimen,

representing the same object of interest from the same cell type. Remember that in this context, the specimen really represents a group of animals, so that a derived data set is the accumulated results typically presented in papers. Because different subsets of the group of animals used in an experiment may be employed for different analysis types, the number of animals was represented in the derived data records table and not in the experiment table. The types of analytic results represented in the database are numbers, distributions, and morphologies. We included a separate table for each of these, with fields for the relevant metadata.

Documentation about the region of interest was captured for each analytic result, and stored in the table called region records. A region record was here defined as a mention of a region of interest in the context of a specific analytic result. The region record table is connected to the tables for the analytic results (number, distribution and morphology tables). The same region record can be relevant for several data sets, and one derived data set might contain data from several regions. An effective many-to-many relationship between these tables was achieved because they are linked through the tables for the analytic results.

To summarise, the Murine basal ganglia database is structured on the following logic. A source can report on one or more experiment. An experiment may give rise to one or several specimens, which are used to generate one or more derived data records. The derived data record can consist of number, distribution or morphology information. For any analytic result, information related to the region of interest and the documentation of it is stored in the region records table. These can be considered to be the main tables in the database, to which information is always entered (although a source might only have one of the analytic result types). Several other tables connected tables contain specialised metadata that is entered if relevant, e.g. about antibodies, light- or electron microscopy, stereology, single-cell visualisation protocols, etcetera.

Comparison to other databases

The Cell Centered Database (CCDB, [41]) contains primary data of the types that are used for the derived data contained in our database (2D and 3D structural data from light and electron microscopy). Thus, a brief comparison of how the database designs relate is of interest. The CCDB is structured around seven main (and required) tables (”Project”, ”Experiment”, ”Subject group”, ”Subject”, ”Tissue/Cell”, ”Processing”, and ”Microscopy product”), with each one having a one-to-many relationship to the next. This is similar to the logic of our database, with four main tables (”Sources”, ”Experiments”, ”Specimen”, ”Derived data

records") connected by one-to-many relationships. In both databases, these tables contain general information about sources, experiments, subjects, and outputs (in CCDB, the main output is the microscopy products, while in our database it is the derived data records). Furthermore, both in the CCDB and our database, there are multiple types of output (their "Tilt Series", "Z series", "Time series", "Mosaic" and "Survey", and our "Quantitations", "Distributions", and "Morphologies") that are stored in separate tables. Lastly, in both cases, detailed specimen preparation and imaging metadata are stored in multiple additional tables.

The CCDB was designed to reflect the process of generating reconstructions from 2D microscopic images, while the Murine basal ganglia database rather was designed around the logic of analysing microscopic images, with some constraints due to how such results are typically reported. For example, because the CCDB has collected primary data and information directly from data producers, their design allows for connecting a microscopy product to the individual subject from which it was obtained (with "Subject group" and "Subject" being two of their main tables). In our database, we did not collect primary data but derived data (analytic results) from the literature, and it is generally not possible to relate numbers to individual subjects used in a study. Information that is likely stored in the CCDB "Subject group" and "Subject" tables is contained in the "Experiments" table of our database. Interestingly, the creators of the CCDB also note that the design of tables capturing details about specimen preparation protocols presented a particular challenge, as such protocols are very flexible [41]. This was also a challenge that quickly became apparent when designing our database.

In summary, the underlying logic of these two databases are relatively similar, with several required tables connected with one-to-many relationships reflecting the chronology of how outputs are generated; specialised metadata for different types of products and tissue processing protocols are then stored in other tables. Of course, the type of metadata stored in the databases will differ at several points since the main outputs are very different; otherwise, differences seem mainly due to the fact that our database had to be designed with the constraints of how information is typically reported in the literature, while the CCDB relies on data providers themselves entering information and data.

3.3.2 Finding sources

To find sources of quantitative information to include in our database, the literature and public repositories were systematically searched. To find public repositories with relevant data, we queried the Neuroscience Information Framwork's database of

resources (https://neuinfo.org/, go to Resources). This database allowed relatively efficient recognition of candidate repositories, which were considered for inclusion. For the literature search, we employed several search strings to query the PubMed database. All the papers returned from these searches were manually screened for relevant data. The search string essentially comprised four parts, with each part specifying 1) the species of interest, 2) the brain regions of interest, 3) the parameters of interest, and 4) the methods of interest. The third part (parameters of interest), specified that a term related to the data of interest would have to appear within five words of the structures of interest:

(((number* or counted or counting or densit* or distribution* or stereolog* or quantita* or quantific*) adj5 (cell* or neuron* or synapse* or bouton* or spine*)).tw,kf.)

The fourth part of the search string (methods of interest) reduced the number of returns considerably:

(Immuno*chem* OR histochem* OR histolog* OR immunofluoresc* OR fluoresc* OR cytochem* OR "electron microscop*".tw,kf. OR ultrastructur* OR cytoarchitect* or chemoarchitect* OR "single-cell label*".tw,kf. OR "intracellular label*".tw,kf. OR "in situ hybridi*".tw,kf. OR (golgi ajd3 stain*).tw,kf.)

However, some studies with relevant data were clearly missed by this approach, probably largely due to this last part of the search string. For example, a very relevant study where neurons were stereologically quantified in all basal ganglia regions [121] was not retrieved. This study included the phrase "number of neurons" in the abstract, which would make it show up using the words related to parameters of interest. However, there is no mention within the abstract of the histological technique that was used. In another relevant paper, examining tyrosine hydroxylase positive cells in the substantia nigra [122], the abstract included several mentions of "neuron number" and similar phrases, but did not mention the immunohistochemical methods. In a third case of a relevant study that was missed in our search [123], the abstract made several mentions of "neuron loss" and "neuron survival". The major focus of that study was elucidating group differences, which then was reflected in the choice of words in the abstract. These three studies provide examples that makes designing efficient search strings difficult. We mitigated some of these challenges in later iterations of the search, by also specifically searching for papers mentioning the word stereology (or lexical variants with the base "stereolog"). However, far more effective literature mining studies could be performed if papers were indexed

according to methodologies used and parameters measured.

3.4 Who's counting and how?

A result of the literature mining was the finding that numbers reported in the literature are highly variable and that cell-type specific data are seldom replicated. In the third paper of this thesis, we therefore quantified the largely non-overlapping parvalbumin and calbindin cell types across the mouse brain. Quantification of various cell types has been a central aim in the field for decades, and a long-standing debate has been *how* cells in the brain should be counted to yield as realistic numbers as possible. In the following, I will discuss the accuracy and efficiency of different methods of counting.

3.4.1 Traditional counting methods

A central challenge when counting cells in 2D sections is that each observed object represents a cell profile. A cell profile seen in a section is not always a unique representation of a cell body: depending on where along the volume of the cell body the cell is cut, it may have profiles appearing in more than one section. Thus, counting profiles directly will lead to falsely high estimates, which has traditionally been solved with mathematical formulas incorporating the size of the cell and the thickness of the section to correct over-counting [124]. Later, stereological procedures were applied to histological material. The central idea with these methods is to combine statistical methods of sampling (systematic random selection of a fraction of the region under study) with the counting of a unique point on each cell (usually the nucleus or nucleolus) [125]. Sampling is performed throughout the thickness of the section, and each unique point is counted only when it comes into focus. Thus, these methods are designed to avoid over-counting in the first place. These two branches of methods have been referred to as model-based (those based on profile counts corrected with mathematical formulas) and design-based (stereological methods) [122] and together constitute the main ways that cells have traditionally been counted.

The debate about which of these approaches should be used to count cells was particularly heated during the late 1990s, when stereological methods were gradually being implemented by an increasing number of researchers. The main argument from those favouring the stereological method was that model-based approaches (e.g. Abercrombie's formula, [124]) would lead to a geometrical bias (i.e. the more inaccurate the size measurements of the cell size, the more geometrical bias

will be present in the resulting estimate, [126]). On the other hand, advocates of traditional methods argued that these biases are minimal as long as certain assumptions are met [127]. Arguments from both sides were usually supported by theoretical examples showing the impact of geometrical bias given various levels of measurement error or for different object-to-section height ratios. In 2001, von Bartheld and colleagues summarized the debate and urged researchers to "stop squabbling over minimal biases when larger sources of error are not even known or understood yet" ([128], page 506).

Because stereological methods are free of the geometrical biases that may apply to formulas designed to correct for over-counting, it is often referred to as "unbiased stereology". However, some of the biases that stereological methods should theoretically avoid may be present when applied to real-world data [128]. Furthermore, many other sources of bias exist than those associated with corrections for over-counting. One particularly under-communicated source of bias is observer bias [129]. It is unrealistic for any method to eliminate observer bias, but clear and transparent documentation of analyses will at least facilitate full interpretation of studies.

3.4.2 Cell segmentation and quantification

As seen above, established counting methods essentially combine manual counting of a fraction of a region with mathematical models [124] or statistical principles [125] to extract estimates of total numbers. Results from such studies are critically dependent on how the region of interest was defined, but researchers often disagree on these definitions. When attempting to integrate data from multiple sources, semantic translation to the closest matching region in a common atlas will necessarily introduce some uncertainty. As we demonstrate in Paper II, there are considerable differences in reported numbers across studies. At least some of this variance is likely to be due to variable definitions of regions of interest. Another concern is that while traditional counting methods may be efficient for quantifying relatively small regions of interest, they are not a feasible alternative for brain-wide analyses.

Over the last decade, interest in semiautomatic or automatic quantification methods in neuroscience has increased. Early automatic segmentation applications were based on simple intensity thresholding, but more sophisticated methods have emerged over time [130]. Segmentation has further gained from advances in machine learning, and the application of active learning methods allows efficient training based on selective human input [131]. Machine learning algorithms has been applied for segmentation and / or quantification of cell types [84, 85, 132–134] and other

features of interest [135–137] across the brain or large regions of it. The three counting methods described here (model-based, stereological, and segmentation based) are illustrated in Figure 6.

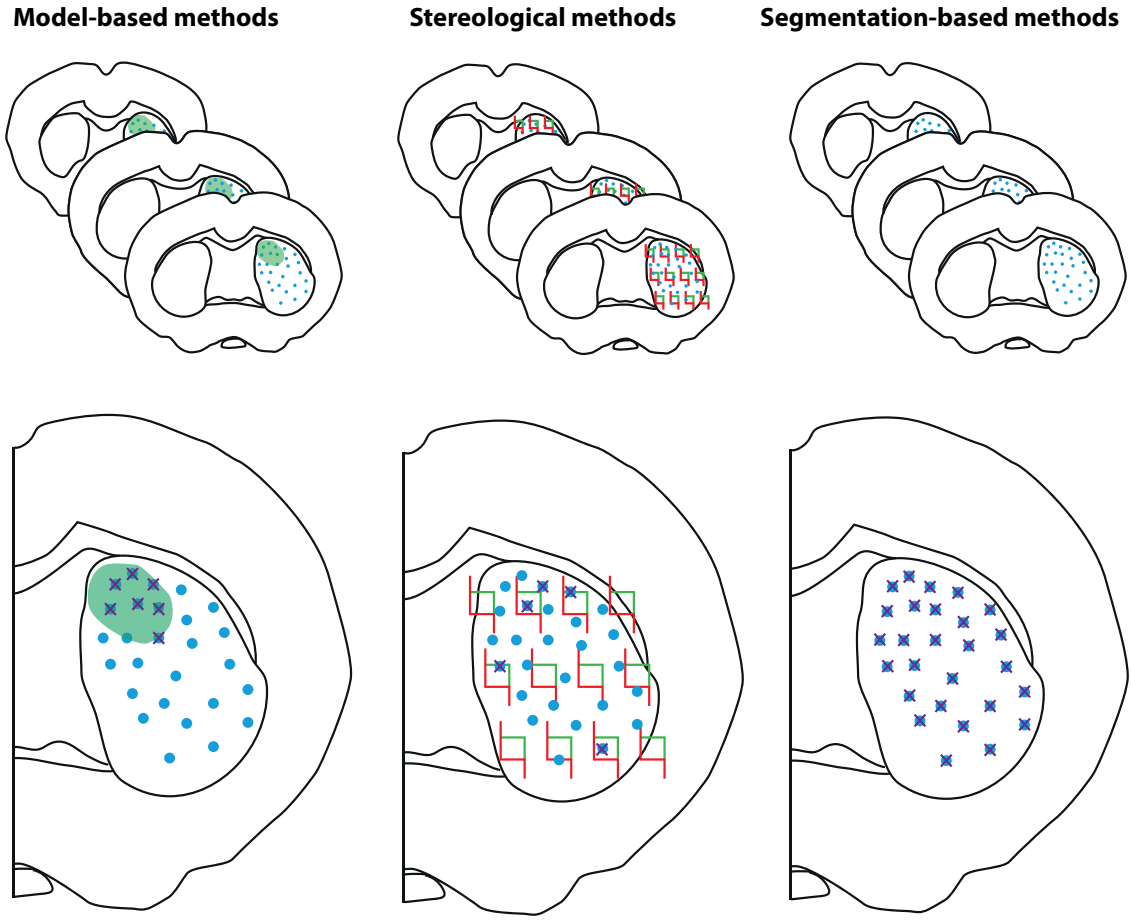


Figure 6. Counting methods used in neuroscience. Illustration of the three counting methods described in the text. Cells are illustrated as blue dots, and the sampled cells for each method are indicated by purple crosses. **Model-based methods** combines counting in small areas (e.g. the dorsomedial striatum, here illustrated in green) with mathematical models (e.g. Abercrombie's formula), to estimate cell densities or total numbers. Because all cells are counted manually, this can typically only be done for small areas of interest. **Stereological methods** employ statistical sampling schemes to place counting frames across a region of interest, and can thus be efficiently applied to larger areas. Counting is performed only within the counting frames, and total number estimates are extrapolated. **Segmentation-based methods** rely on automatic or semi-automatic recognition of features of interest in digital images, which ideally allows sampling of all cells across large areas.

Efficient approaches to counting cells will be essential for analytic efforts to scale to the rate of data production in the field. However, most of the published work in this direction is based on custom-written codes that (if shared) require significant

skill to use and are often essentially tailored to the specific use-case in question (or applies only to certain types of material, e.g. fluorescently stained mouse brain sections). To meet this challenge, Yates and colleagues [137] recently published the QUINT workflow, which combines three open-access, GUI-based softwares for image segmentation, atlas registration and quantification. In addition to being easily applicable for neuroscientists who do not possess advanced coding skills, the QUINT workflow incorporates brain reference atlases for both the rat and mouse. In the third paper, we used this workflow to quantify calbindin and parvalbumin neurons in immunohistochemically stained material from rat and mouse brains. We implemented a series of post-processing steps, including use of Abercombie's formula [124], to correct numbers generated by the QUINT workflow and extract absolute number and volumetric density estimates.

As is clear from the sections above, choosing how to count cells is not a trivial question with a single ideal solution that will fit every researcher and every purpose. Stereological methods may provide the most realistic estimates for 2D material, but requires measurements to be made throughout the section thickness. This is essentially not possible for published, digital image data. In these cases, traditional 2D counting approaches with double-counting corrections [124] provide a more pragmatic alternative. Such methods have been shown to generate realistic numbers [122], and can also advantage from advances in image segmentation to increase efficiency and re-usability of analyses. Regardless of the method used, results will depend on both the parameters of the analytic workflow (how are objects counted?) and on the observer (who is counting objects?). In order for results to be truly possible to interpret and reproduce by other researchers, focus should be on providing full access to original material *and* analytic results.

Chapter 4

Discussion

The primary aim addressed in this thesis was to establish new standards and approaches for mining, collecting, integrating, and sharing diverse neuroscience data from the literature, image repositories, and novel experimental work into common reference atlases of the rat and mouse brain. Detailed methodological considerations have been elaborated in the discussion section above, and I will give a more general discussion in the following. I will first discuss current practices for reporting anatomical metadata and the use of atlases in the field. I will also discuss how ambiguities in descriptions of anatomical locations can be resolved and data integrated by new reporting standards and methods for semantic and spatial registration to atlas. Secondly, I will discuss the coverage that the Murine basal ganglia database has of data available from the literature and the utility of this data collection. Lastly, I will give some perspectives on reporting and sharing of data in neuroscience - some general ones and some specifically related to cell counting data.

4.1 Reporting and integrating data through atlases

It is apparent from both the first and second paper of this thesis that the use of neuroanatomical terms are highly variable and that current practices for reporting anatomical metadata are often insufficient to support interpretation, replication, and re-use of data. To mitigate this situation, clear reporting standards and solutions for integration of both new and existing data are needed. To meet this need, the work described in this thesis has involved 1) establishing minimum and best practice recommendations for reporting anatomical metadata; 2) demonstrating the use of spatial registration to atlas as an alternative for a range of neuroscientific methods; 3) establishing and implementing a workflow for comparing and translating terms found in different reference atlases and custom terms; and 4) demonstrating the utility of 3D atlas registration in efficient and transparent brain-wide analysis of cell numbers.

4.1.1 The changing landscape of neuroanatomy

Our work on comparing different atlases demonstrated that nomenclatures vary considerably across atlases and even versions of the same atlas. Some terms, like "caudoputamen" and "subthalamic nucleus", are heavily used across atlases, but for regions that are typically subdivided (e.g. substantia nigra and globus pallidus), several different terms and lexical variants are used depending on the atlas or atlas version. Changes between atlas versions typically reflect new knowledge or changes in views over time, but definitions also differ depending on *who* is making the atlas and the criteria it is based on. These inter-atlas differences are far more pronounced than differences between atlases for rats and mice. Thus, Paxinos and Watson's reference atlases of the rat brain are more similar to the mouse brain atlases by Franklin and Paxinos than it is to the rat brain atlases by Swanson. Swanson's rat brain atlases, on the other hand, have features that are very similar to the Allen Mouse Brain atlas [138] (the latter was made based on the nomenclature of the former). With rapidly changing atlases, unambiguous reporting of anatomical regions in the literature becomes all the more important.

However, the use of terms in the field does not necessarily reflect the use of terms in atlases. As described in Paper II, we found less than half of the terms used in the literature to be consistent with any reference atlas. In some cases, completely different terms may be used in the literature than in atlases. The caudoputamen is one example of this: the term appears consistently across all atlases investigated

here, but it is equally often referred to as the "striatum" in the literature. Because the striatum is also commonly used to refer to the combined caudoputamen and nucleus accumbens, this practice creates significant confusion when interpreting findings. Interestingly, the term "striatum" is often used in combination with reference to an atlas where the term does not appear, even when the data are clearly from the caudoputamen part. Another difference is that while atlases are typically very detailed with respect to subregions (e.g. "accumbens nucleus, rostral pole" and "substantia nigra, compact part, dorsal tier"), these terms are seldom used in the literature. Instead, collective terms like "nucleus accumbens" and "substantia nigra" are very commonly used. Conversely, for large regions that are not commonly subdivided (e.g. the caudoputamen), specification of zones is common, such as "Dorsomedial striatum". Such zones are typically not distinguishable based on defined criteria, and seem to be used mainly to indicate the rough position within a larger region of interest. It has been noted before that while several efforts have been made to create standardised terminologies for neuroanatomy [139], their use in research reporting is rather limited. That is, French and colleagues found that 77 % of neuroanatomical terms found in terminologies were not found in a selection of more than 12 000 abstracts from the Journal of Comparative Neurology [140]. We noticed that among the papers that did use atlases, the earliest versions of the rat and mouse brain atlases by Paxinos and colleagues (published in the 1980's and 90's) were most often cited. This was true for even very recent publications. Increasing focus on providing open access to atlas resources [32, 67, 78] will hopefully enable researchers to use more updated versions of atlases.

To facilitate unambiguous interpretation of studies, we defined minimum and best practice recommendations for documenting anatomical locations. In this process, we identified nine factors (see "Methodological considerations" for details) that we believe together cover the various ways that neuroanatomical locations may be documented. This is supported by the fact that throughout the survey, we did not encounter any form of anatomical metadata that did not fit in this scheme. However, some forms of documentation (such as information about spatial registration to a reference atlas) were very rarely seen. These nine factors were used as the basis for our recommendations in Paper I, and further used in Paper II to generate, for each region of interest mentioned in the literature survey, a "Documentation score" ranging from 0 to 10. Such a scoring system might be relevant for ongoing efforts to create metrics of rigor and transparency in research [141]. While it is unlikely that researchers will adopt a single standard nomenclature for anatomical regions within the near future, or indeed ever, standards for how to document regions can at least facilitate unambiguous interpretation of studies.

4.1.2 Navigating murine brain atlases

For anatomical metadata to be used to integrate data, they need to follow a single consistent standard terminology or be linked to a common coordinate-based framework. Given the fact that many different atlases exist and that they are constantly evolving with respect to nomenclature and definitions of regions, the ability to navigate between atlases - i.e., atlas interoperability - will be of importance as well. That is, it should be possible to make translations across both terminologies and coordinate-systems of two atlases. The strategies described throughout the work in this thesis, including spatial registration to common atlases [54] and semantic translation across atlases, facilitate atlas interoperability.

The spatial registration strategies demonstrated in this thesis ensures that data are related to the template of an atlas, i.e. the brain used to create the atlas. As long as this template does not change, spatially registered data can be re-used with alternative delineation and nomenclature versions. Interoperability between atlases using different templates is slightly more challenging. For this to be feasible, the brain atlases in question would need to be interconnected via a single common coordinate system. A digital atlasing infrastructure addressing this challenge was described by Zaslavsky and colleagues [142], where the openly accessible Waxholm Space for the Sprague-Dawley rat brain [78] and C57BL6/J mouse brain [143] provides the spatial anchors for other atlases. Such solutions [142] will be essential to relate data across resources. Our spatial co-registration data sets may be useful to define the parameters needed to use the Waxholm Space rat brain atlas and Allen Mouse Brain Common Coordinate Framework as spatial anchors for classical stereotaxic atlases [55–57, 60–67, 96] and data connected to them.

The semantic registration workflow through translation of terms developed here constitutes a solution for location integration of published data, effectively making atlas terminologies as interoperable as possible. The metadata collected for each semantic registration are also important for location integration purposes. We implemented a detailed evaluation of the definition of regions of interest reported in studies, and how this relates to other ways of defining regions. We chose to use the atlases used in EBRAINS [32, 78–80] as our references, and all the derived data registered in our database were semantically registered to the relevant EBRAINS atlas. We evaluated how comparable each region of interest mentioned in the literature was to their corresponding region in the relevant EBRAINS atlas. In this process, we recorded several new types of metadata that together give information about whether the whole region has been sampled and if it has been specifically

sampled, as well as the documentation available to support the confidence of the registration. These metadata should be of more general relevance for semantic registration to atlas, which is an important part of the atlas curation workflow in EBRAINS.

4.2 Utility of the Murine basal ganglia database

A common problem across several neuroscience database resources, as discussed by Ascoli [144], is lack of coverage of existing data (i.e. "information sparseness"). While our database, presented in Paper II, constitutes the largest collection of quantitative neuroanatomical information from the basal ganglia to date, there is certainly more information existing in the literature than is available through the database. For example, we excluded some 150 papers in which numbers were available in graphs only. Additionally, we only extended the search to include all control animals of all ages for the substantia nigra and the caudoputamen. It was not our intention to cover all relevant quantitative data that exists in various formats across the community. Rather, our aim was to establish workflows through which such quantitative data can be unified, organized and shared. The quantitative database of the basal ganglia was as such intended to provide a proof-of-principle of the novel workflows, and to illustrate methodological and analytical steps as well as reporting practices that may contribute to data inconsistencies. The crucial point made by Ascoli [144] is that databases become increasingly useful as they are densely populated. This point was re-iterated by Akil and colleagues [145], who argue that adding information to existing databases will be an important alternative to creating new ones. We openly share the entire database in our Paper II, and additionally describe a workflow through which other researchers can collect more data, integrate it in the database, and share it with the community. Thus, others may build on this effort to increase coverage. While this workflow is currently supported by custom-made MS Excel sheets and MS Access database, future developments may streamline these processes, so that estimates can be directly entered to the EBRAINS Knowledge Graph. Our efforts can guide such developments, in that it provides specifications of relevant metadata and conversions that should be supported.

Our database revealed large variations in estimates across studies. Thus, dense coverage of data that are not comparable due to lack of rigor in either the experimental or reporting practices (or both), would probably not be useful, nor would it contribute to less confusion for the user. In other words, it might not be constructive to cover large amounts of data that cannot be considered reproducible

or reliable. This is not necessarily a simple question of which method has been used, but rather depends on the rigor with which the method has been performed and documented. A neuronal number estimate obtained by a model-based method (e.g. Abercrombie's formula [124]) using appropriate corrections might be more valid than a similar estimate obtained by a poorly performed study using a more modern stereological approach [127, 146, 147]. Indeed, in a recent study comparing both stereological estimation and Abercrombie's method to numbers obtained by serial reconstruction of the cell population, both procedures were found to yield realistic numbers [122]. Considering one method to be a gold standard that is always preferred over another is thus not enough, since all methods have assumptions that must be recognized and met. Verifying that the assumptions of a method has been met, although a more complex task than simply considering what method was used, is thus an important part of considering the quality of the data. As our results show, such verification is often not possible for the published literature due to lack of critical metadata. Thus, given the manual effort required to find appropriate data and organize it through the pipelines outlined here, future efforts should focus primarily on data considered to be of high quality. One suggestion would be to require that data are (1) related to a clearly defined region of interest (perhaps even requiring full coverage, at least for total number estimates) and (2) obtained by a counting method where error is minimised and assumptions can be verified. Having a set of criteria would not only limit the scope and time requirements for manual curation efforts, but give the potential for denser coverage of curated quantitative data.

It has been shown that even sparse quantitative and morphological data can contribute to realistic, dense computational models [148]. The current version of the database, although somewhat sparsely populated, should therefore provide useful estimates for computational modelling efforts. In addition to providing easy access to a number of parameters, the papers that were excluded are listed with the reason for exclusion indicated. This list constitutes an easy source of references to extend the current database. The numbers present in the database might also be combined with ratios provided in the literature to provide tentative estimates for parameters that has yet to be experimentally quantified. Such an approach has been used previously Bezaire and Soltesz [149], who through literature mining and careful calculations obtained biological constraints for a computational model of the hippocampus CA1. A cycle where existing data is used in computational models to predict values that can later be experimentally tested is a key aspect of computational modelling [14, 150].

4.3 Documenting and sharing neuroscience data

4.3.1 Reporting standards

A central principle of the scientific method is that findings reported in a paper should be possible to reproduce. Since this would require a thorough understanding of all the methodological conditions under which they were obtained, reproducibility is to a high degree a question of the transparency and completeness of scientific reports. Publishers therefore typically require researchers to report methodologies in such a way that another researcher could replicate the study. At the same time, concise text is often encouraged. Thus, in reality, researchers do not report every detail of an experimental setup in a journal article.

As seen in both Paper I and II, the methodological details reported in journal articles is more often than not insufficient to interpret and compare results across studies. Thus, statements that methodology sections should provide the necessary and sufficient details to replicate a study has not led to satisfactory reporting practices in the field. To meet this challenge, metadata standards are being developed across multiple subfields [24, 151, 152]. Furthermore, the Research Resource Identifiers (RRIDs; [153]) ensures that software, antibodies, cell lines, and animal lines are assigned unique identifiers that may be searched through a single portal. This initiative has led to significant progress on the reporting practices for these types of resources [23]. From the publishers' side, many journals no longer have restrictions on the length of methods sections. The Cell Press journals have furthermore implemented "STAR Methods" sections [154] where standardized sections and tables are used to ensure that both used resources and procedures are listed. Such efforts can go a long way to improve reporting, but data sharing practices will be important to facilitate sharing of all the elements needed to understand and replicate a study, and not least to re-use data.

4.3.2 Challenges and progress for neuroscience data sharing

Despite emphasis on the value of sharing data over the last decades, this has yet to become the norm in the field of neuroscience for several reasons, including technical, legal, conceptual and cultural ones. Many of these have been articulated for a long time [7]. Significant progress has been made over the last two decades in terms of web technology in general [155] and storage and computing in particular [156] that facilitate easy exchange of data. There has also progress towards the use of clear policies and licences to govern how data can be re-used. When data from

human subjects are to be shared, ensuring anonymization and adherence to privacy regulations is a non-trivial task [11, 155].

Another important challenge is the heterogeneity of neuroscience data. It is not immediately obvious how complex and diverse neuroscience data sets can be organized in a meaningful way. How should information about an electrophysiological recording be combined with data from an immunohistochemistry experiment? What type of information will be needed to interpret these studies? Which pieces of information should be available for all neuroscience data sets, and which are specifically relevant to certain methods? The extent to which data sets are well-organised and tagged with relevant information impacts the ease with which shared data can be found and retrieved, ultimately affecting the perceived usefulness of sharing data. As shown throughout this thesis, anatomical metadata can provide an important basis for integrating neuroscience data across modalities.

But perhaps the most prevailing obstacle to sharing data in neuroscience has been cultural. There are several reasons why neuroscientists may be hesitant to share their primary data: reluctance to let others access hard-won data or a fear of losing a competitive edge, the time required to organize and upload data, or worry that others might find mistakes or criticise the data [5, 144]. There has also been a lack of clear incentives to share. As argued by Ascoli [157], there should be a balance between the effort required to share data and the gains for the data provider. While there are examples where data sharing has led to important new discoveries [158] and indications that data articles have higher impact than traditional ones [159], it seems that the advantages of sharing primary data have so far not outweighed actual or perceived costs for neuroscience researchers.

External pressure to share data is growing across scientific fields. For example, several publishers now have explicit and detailed data sharing policies, e.g. PLOS [160], Nature (www.springernature.com/gp/authors/research-data-policy), and Science (www.sciencemag.org/authors/, go to Editorial policies). Funders are also implementing policies on open access to research data, for example by promoting the inclusion of data management plans in grant applications (https://www.forskningsradet.no/en/Adviser-research-policy/open-science/open-access-to-research-data/). Generally, policies from publishers mandate that a statement about how the data are shared is included in the paper, but do not enforce specific sharing practices *per se*. Analyses of data availability statements from different journals have indicated that less than 20 % of articles typically link to data available from a repository, the recommended way to share data [161, 162]. A stricter policy might improve this situation. However, it is important to note that data sharing mandates like these do not take into account incentives for the primary

data generators. As some have argued, "compliance-focused sharing" might simply result in depositions of data sets that are virtually impossible to find, interpret, or replicate [163]. Since shared data are most useful once they are well described and put in context of other data, the value of sharing data is dependent on the available solutions for their organisation and annotation, as discussed above. Thus, the value of data sharing policies depends on how they are enforced, and both peer review practices and evaluation by funding agencies may need to be adjusted to validate that data sharing statements or plans are satisfactory. Ensuring that researchers get credit for the effort that is put into sharing valuable data collections will also be essential [164].

4.3.3 FAIR cell counting data

Our results show that published cell counting data have limited comparability and that cell-type specific information is typically only available for a few regions and with few replications across studies. Given the wide-spread interest in such data, there is a need for efficient quantitative analyses of different cell types on a brain-wide scale, but traditional methods do not easily scale to meet this need. Furthermore, re-use of analyses is not a realistic ideal when considering traditional counting methods, where results are usually obtained by manual counting of selected sampling sites and in some cases dependent on sampling throughout the section thickness [125] (thus requiring access to the physical raw data). Advances in machine learning has made automatic and semi-automatic quantification pipelines a feasible alternative to traditionally manual tasks [165], and in theory allows for sharing analytic results in a way that would allow re-use of all generations of the data. However, to date sharing practices that allow for re-running analyses with alternative regions of interest or segmentation algorithms has not been in place. In response to these challenges, we used the QUINT workflow [86] to analyse brain-wide data sets from both rats and mice, showing parvalbumin (rat and mouse) and calbindin (mouse) cells, and corrected the resulting numbers using Abercrombie's formula [124]. We furthermore shared all parts of the data used for analysis. The QUINT workflow flexibly combines software for different steps of the process (image segmentation, atlas registration and quantification), allowing re-use of selected parts of an analysis. Our analysis of these data sets and sharing of all the generations of the data (raw, processed raw, derived) as well as detailed metadata about the procedures, provides a demonstration of how cell counting can be performed in a transparent and re-usable way. Thus, both primary and derived data are Findable, Accessible, Interoperable and Re-usable in accordance with the FAIR principles [51]. Going forwards, different subfields of neuroscience will need to converge on reporting and sharing practices that facilitate

re-use of specific types of data. Our work with cell counting data provides a good starting point for defining such practices for quantitative neuroanatomical data.

4.4 Conclusion

The primary contributions of this thesis to the field of neuroscience and neuroinformatics are new standards for reporting anatomical metadata; new workflows for semantic and spatial atlas registration; and demonstration of how spatial registration can be used to perform brain-wide analyses. More specificially, through the work presented in this thesis, we have:

1. Established minimum and best practice recommendations for determining and documenting anatomical location, and demonstrated that spatial registration to reference atlas is a feasible alternative across a range of neuroscientific methods
2. Created the most comprehensive collection to date of quantitative cellular measurements from the basal ganglia, and shown that numbers are highly variable and that cell-type specific information is rarely replicated
3. Performed brain-wide analysis of defined cell types in rat and mouse brains, and demonstrated how these data can be reported transparently and to facilitate re-use

Thus, in addition to providing quantitative information about several parameters across the basal ganglia regions and selected cell types across the brain, the papers and data sets published through the work with this thesis provide important demonstrations of how digital atlases can be used to collect, organise and integrate data and how neuroscience data can be shared to facilitate their re-use.

References

[1] Y Frégnac. "Big data and the industrialization of neuroscience: A safe roadmap for understanding the brain?" In: *Science* 358.6362 (2017), pp. 470–477. DOI: 10.1126/science.aan8866.

[2] ME Martone, A Gupta, and MH Ellisman. "e-Neuroscience: Challenges and triumphs in integrating distributed data from molecules to brains". In: *Nature Neuroscience* 7.5 (2004), pp. 467–472. DOI: 10.1038/nn1229.

[3] J-B Poline, JJ Chodacki, and A Gulick. "From data sharing to data publishing". In: *MNI Open Research* 2.1 (2019), pp. 1–12. DOI: 10.12688/mniopenres.12772.2.

[4] EC McKiernan et al. "How open science helps researchers succeed". In: *eLife* 5.e16800 (2016), pp. 1–19. DOI: 10.7554/eLife.16800.

[5] S H Koslow and S Subramaniam, eds. *Databasing the brain: From data to knowledge*. New York: John Wiley & Sons, Inc., 2005.

[6] TJ Sejnowski, PS Churchland, and JA Movshon. "Putting big data to good use in neuroscience". In: *Nature Neuroscience* 17.11 (2014), pp. 1440–1441. DOI: 10.1038/nn.3839.

[7] S-I Amari et al. "Neuroinformatics: the integration of shared databases and tools towards integrative neuroscience." In: *Journal of Integrative Neuroscience* 1.2 (2002), pp. 117–128. DOI: 10.1142/S0219635202000128.

[8] JG Bjaalie and S Grillner. "Global Neuroinformatics: The International Neuroinformatics Coordinating Facility". In: *Journal of Neuroscience* 27.14 (2007), pp. 3613–3615. DOI: 10.1523/JNEUROSCI.0558-07.2007.

[9] JG Bjaalie. "Understanding the brain through neuroinformatics". In: *Frontiers in Neuroscience* 2.1 (2008), pp. 19–21. DOI: 10.3389/neuro.01.022.2008.

[10] LA Jorgenson et al. "The BRAIN initiative: Developing technology to catalyse neuroscience discovery". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 370.1668 (2015). DOI: 10.1098/rstb.2014.0164.

[11] AL Vaccarino et al. "Brain-CODE: A secure neuroinformatics platform for management, federation, sharing and analysis of multi-dimensional neuroscience data". In: *Frontiers in Neuroinformatics* 12.28 (2018), pp. 1–14. DOI: 10.3389/fninf.2018.00028.

[12] H Okano et al. “Brain/MINDS: A Japanese national brain project for marmoset neuroscience”. In: *Neuron* 92.3 (2016), pp. 582–590. DOI: 10.1016/j.neuron.2016.10.018.

[13] K Amunts et al. “The Human Brain Project—Synergy between neuroscience, computing, informatics, and brain-inspired technologies”. In: *PLoS Biology* 17.7 (2019), e3000344. DOI: 10.1371/journal.pbio.3000344.

[14] H Markram et al. “Reconstruction and simulation of neocortical microcircuitry”. In: *Cell* 163.2 (2015), pp. 456–492. DOI: 10.1016/j.cell.2015.09.029.

[15] J Olesen and M Leonardi. “The burden of brain diseases in Europe”. In: *European Journal of Neurology* 10.5 (Sept. 2003), pp. 471–7. DOI: 10.1046/j.1468-1331.2003.00682.x.

[16] L Bornmann and R Mutz. “Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references”. In: *Journal of the Association for Information Science and Technology* 66.11 (2015), pp. 2215–2222. DOI: 10.1002/asi.23329.

[17] JPA Ioannidis. “Why most published research findings are false”. In: *PLoS Medicine* 2.8 (2005), pp. 696–701. DOI: 10.1371/journal.pmed.0020124.

[18] M Bota, H-W Dong, and LW Swanson. “Brain Architecture Management System”. In: *Neuroinformatics* 3.1 (2005), pp. 15–48. DOI: 10.1385/NI:3:1:015.

[19] B Horwitz. “Integrating neuroscientific data across spatiotemporal scales”. In: *Comptes Rendus - Biologies* 328.2 (2005), pp. 109–118. DOI: 10.1016/j.crvi.2004.10.015.

[20] GA Ascoli. “From data to knowledge”. In: *Neuroinformatics* 1.2 (2003), pp. 145–147. DOI: 10.1007/s12021-003-0001-2.

[21] SL Hill. “How do we know what we know? Discovering neuroscience data sets through minimal metadata”. In: *Nature Reviews Neuroscience* 17.12 (2016), pp. 735–736. DOI: 10.1038/nrn.2016.134.

[22] NA Vasilevsky et al. “On the reproducibility of science: unique identification of research resources in the biomedical literature”. In: *PeerJ* 1 (2013), e148. DOI: 10.7717/peerj.148.

[23] A Bandrowski et al. “The Resource Identification Initiative: A cultural shift in publishing”. In: *Neuroinformatics* 14.2 (2016), pp. 169–182. DOI: 10.1007/s12021-015-9284-3.

[24] EW Deutsch et al. “Minimum information specification for in situ hybridization and immunohistochemistry experiments (MISFISHIE).” In: *Nature Biotechnology* 26.3 (2008), pp. 305–312. DOI: 10.1038/nbt1391.

[25] S Polavaram and GA Ascoli. “An ontology-based search engine for digital reconstructions of neuronal morphology”. In: *Brain Informatics* 4.2 (2017), pp. 123–134. DOI: 10.1007/s40708-017-0062-x.

[26] FT Imam et al. “Development and use of ontologies inside the Neuroscience Information Framework: A practical approach”. In: *Frontiers in Genetics* 3.111 (2012), pp. 1–12. DOI: 10.3389/fgene.2012.00111.

[27] LW Swanson. "What is the brain?" In: *Trends in Neurosciences* 23.11 (2000), pp. 519–527. DOI: 10.1016/S0166-2236(00)01639-8.

[28] G Ascoli et al. "Petilla terminology: nomenclature of features of GABAergic interneurons of the cerebral cortex." In: *Nature Reviews Neuroscience* 9.7 (2008), pp. 557–568. DOI: 10.1038/nrn2402.

[29] DJ Hamilton et al. "An ontological approach to describing neurons and their relationships". In: *Frontiers in Neuroinformatics* 6.15 (2012), pp. 1–11. DOI: 10.3389/fninf.2012.00015.

[30] SM Sunkin et al. "Allen Brain Atlas: an integrated spatio-temporal portal for exploring the central nervous system". In: *Nucleic Acids Research* 41 (2012), pp. 996–1008. DOI: 10.1093/nar/gks1042.

[31] ES Lein et al. "Genome-wide atlas of gene expression in the adult mouse brain". In: *Nature* 445.7124 (2007), pp. 168–176. DOI: 10.1038/nature05453.

[32] SW Oh et al. "A mesoscale connectome of the mouse brain". In: *Nature* 508.7495 (2014), pp. 207–214. DOI: 10.1038/nature13186.

[33] B Tasic et al. "Adult mouse cortical cell taxonomy revealed by single cell transcriptomics". In: *Nature Neuroscience* 19.2 (2016), pp. 335–46. DOI: 10.1038/nn.4216.

[34] RA McDougal et al. *Twenty years of ModelDB and beyond: Building essential modeling tools for the future of neuroscience*. 2017. DOI: 10.1007/s10827-016-0623-7.

[35] MN Economo et al. "A platform for brain-wide imaging and reconstruction of individual neurons". In: *eLife* 5.e10566 (2016), pp. 1–22. DOI: 10.7554/eLife.10566.

[36] GA Ascoli, DE Donohue, and M Halavi. "NeuroMorpho.Org: A central resource for neuronal morphologies". In: *Journal of Neuroscience* 27.35 (2007), pp. 9247–9251. DOI: 10.1523/JNEUROSCI.2055-07.2007.

[37] H Hintiryan et al. "The mouse cortico-striatal projectome". In: *Nature Neuroscience* 19.8 (2016). DOI: 10.1038/nn.4332.

[38] B Zingg et al. "Neural networks of the mouse neocortex". In: *Cell* 156.5 (2014), pp. 1096–1111. DOI: 10.1016/j.cell.2014.02.023.

[39] JW Bohland et al. "A proposal for a coordinated effort for the determination of brainwide neuroanatomical connectivity in model organisms at a mesoscopic scale". In: *PLoS Computational Biology* 5.3 (2009). DOI: 10.1371/journal.pcbi.1000334.

[40] M Jeong et al. "Comparative three-dimensional connectome map of motor cortical projections in the mouse brain." In: *Scientific Reports* 6.20072 (2016), pp. 1–14. DOI: 10.1038/srep20072.

[41] ME Martone et al. "The Cell Centered Database project: An update on building community resources for managing and sharing 3D imaging data". In: *Journal of Structural Biology* 161.3 (2008), pp. 220–231. DOI: 10.1016/j.jsb.2007.10.003.

[42] R Bakker, T Wachtler, and M Diesmann. "CoCoMac 2.0 and the future of tract-tracing databases". In: *Frontiers in Neuroinformatics* 6.30 (2012), pp. 1–6. DOI: 10.3389/fninf.2012.00030.

[43] N M van Strien, N L Cappaert, and M P Witter. "The anatomy of memory: an interactive overview of the parahippocampal-hippocampal network". In: *Nat Rev Neurosci* 10.4 (2009), pp. 272–282. DOI: 10.1038/nrn2614.

[44] J Sugar et al. "The retrosplenial cortex: intrinsic connectivity and connections with the (para)hippocampal region in the rat. An interactive connectome". In: *Frontiers in Neuroinformatics* 5.7 (2011), pp. 1–13. DOI: 10.3389/fninf.2011.00007.

[45] SJ Tripathy et al. "NeuroElectro: a window to the world's neuron electrophysiology data". In: *Frontiers in Neuroinformatics* 8.40 (2014), pp. 1–11. DOI: 10.3389/fninf.2014.00040.

[46] DW Wheeler et al. "Hippocampome.org: a knowledge base of neuron types in the rodent hippocampus". In: *eLife* 4 (2015), pp. 1138–1142. DOI: 10.7554/eLife.09960.

[47] D Keller, C Erö, and H Markram. "Cell densities in the mouse brain: A systematic review". In: *Frontiers in Neuroanatomy* 12.83 (2018), pp. 1–21. DOI: 10.3389/fnana.2018.00083.

[48] MS Bienkowski et al. "Integration of gene expression and brain-wide connectivity reveals the multiscale organization of mouse hippocampal networks". In: *Nature Neuroscience* 21.11 (2018), pp. 1628–1643. DOI: 10.1038/s41593-018-0241-y.

[49] Y Ko et al. "Cell type-specific genes show striking and distinct patterns of spatial expression in the mouse brain." In: *Proceedings of the National Academy of Sciences of the United States of America* 110.8 (2013), pp. 3095–100. DOI: 10.1073/pnas.1222897110.

[50] S Polavaram et al. "Statistical analysis and data mining of digital reconstructions of dendritic morphologies." In: *Frontiers in Neuroanatomy* 8.138 (2014), pp. 1–16. DOI: 10.3389/fnana.2014.00138.

[51] MD Wilkinson et al. "The FAIR Guiding Principles for scientific data management and stewardship." In: *Scientific Data* 3 (2016), pp. 1–9. DOI: 10.1038/sdata.2016.18.

[52] JG Bjaalie. "Localization in the brain: New solutions emerging". In: *Nature Reviews Neuroscience* 3.4 (2002), pp. 322–325. DOI: 10.1038/nrn790.

[53] J Boline, E-F Lee, and AW Toga. "Digital atlases as a framework for data sharing". In: *Frontiers in Neuroscience* 2.1 (2008), pp. 100–106. DOI: 10.3389/neuro.01.012.2008.

[54] IE Bjerke et al. "Data integration through brain atlasing: Human Brain Project tools and strategies." In: *European Psychiatry* 50 (2018), pp. 70–76. DOI: 10.1016/j.eurpsy.2018.02.004.

[55] G Paxinos and C Watson. *The rat brain in stereotaxic coordinates*. 2nd ed. Academic Press, 1986.

[56] G Paxinos and C Watson. *The rat brain in stereotaxic coordinates*. 4th ed. San Diego, CA: Academic Press, 1998.

[57] G Paxinos and C Watson. *The rat brain in stereotaxic coordinates*. 5th ed. San Diego, CA: Elsevier, 2005.

[58] G Paxinos et al. *Chemoarchitectonic atlas of the rat brain*. 2nd ed. San Diego, CA: Academic Press Inc, 2008.

[59] G. Paxinos and C. Watson. *The rat brain in stereotaxic coordinates*. 7th ed. Burlington, NJ: Elsevier Inc, 2013.

[60] G Paxinos and C Watson. *The rat brain in stereotaxic coordinates*. 3rd ed. San Diego, CA: Academic Press, 1997.

[61] G Paxinos and KBJ Franklin. *The mouse brain in stereotaxic coordinates*. 2nd ed. San Diego, CA: Academic Press, 2001.

[62] G Paxinos and C Watson. *The rat brain in stereotaxic coordinates*. 6th ed. Burlington, MA: Academic Press, 2007.

[63] G Paxinos and KBJ Franklin. *The mouse brain in stereotaxic coordinates*. 4th ed. San Diego, CA: Academic Press, 2012.

[64] LW Swanson. *Brain Maps: Structure of the rat brain*. 1st ed. Elsevier, 1992.

[65] LW Swanson. *Brain Maps II: Structure of the rat brain*. 2nd ed. Elsevier, 1998.

[66] LW Swanson. *Brain Maps III: Structure of the rat brain*. 3rd ed. Elsevier, 2004, p. 215.

[67] LW Swanson. "Brain maps 4.0—Structure of the rat brain: An open access atlas with global nervous system nomenclature ontology and flatmaps". In: *Journal of Comparative Neurology* 526.6 (2018), pp. 935–943. DOI: 10.1002/cne.24381.

[68] Y Ma et al. "A three-dimensional digital atlas database of the adult C57BL/6J mouse brain by magnetic resonance microscopy". In: *Neuroscience* 135.4 (2005), pp. 1203–1215. DOI: 10.1016/j.neuroscience.2005.07.014.

[69] N Kovačević et al. "A three-dimensional MRI atlas of the mouse brain with estimates of the average and variability". In: *Cerebral Cortex* 15.5 (2005), pp. 639–645. DOI: 10.1093/cercor/bhh165.

[70] T Hjornevik et al. "Three-dimensional atlas system for mouse and rat brain imaging data". In: *Frontiers in Neuroinformatics* 1.4 (2007), pp. 1–11. DOI: 10.3389/neuro.11.004.2007.

[71] H Ni et al. "A robust image registration interface for large volume brain atlas". In: *Scientific Reports* 10.1 (2020), pp. 1–16. DOI: 10.1038/s41598-020-59042-y.

[72] IE Bjerke et al. "Navigating the murine brain: Toward best practices for determining and documenting neuroanatomical locations in experimental studies". In: *Frontiers in Neuroanatomy* 12.82 (2018), pp. 1–15. DOI: 10.3389/FNANA.2018.00082.

[73] MA Puchades et al. “Spatial registration of serial microscopic brain images to three-dimensional reference atlases with the QuickNII tool”. In: *PLOS ONE* 14.5 (2019). DOI: 10.1371/journal.pone.0216796.

[74] DM Simmons and LW Swanson. “Comparing histological data from different brains: Sources of error and strategies for minimizing them”. In: *Brain Research Reviews* 60.2 (2009), pp. 349–367. DOI: 10.1016/j.brainresrev.2009.02.002.

[75] M Hawrylycz et al. “Digital atlasing and standardization in the mouse brain.” In: *PLoS Computational Biology* 7.2 (2011), pp. 1–6. DOI: 10.1371/journal.pcbi.1001065.

[76] L Ng et al. “The INCF Digital Atlasing Program: Report on Digital Atlasing Standards in the Rodent Brain”. In: *Nature Precedings* (2009). DOI: 10.1038/npre.2009.4000.1.

[77] ML Johnson et al. “Androgen Decreases Dopamine Neurone Survival in Rat Midbrain”. In: *Journal of Neuroendocrinology* 22.4 (2010), pp. 238–247. DOI: 10.1111/j.1365-2826.2010.01965.x.

[78] EA Papp et al. “Waxholm Space atlas of the Sprague Dawley rat brain”. In: *NeuroImage* 97 (2014), pp. 374–386. DOI: 10.1016/j.neuroimage.2014.04.001.

[79] LJ Kjonigsen et al. “Waxholm Space atlas of the rat brain hippocampal region: Three-dimensional delineations based on magnetic resonance and diffusion tensor imaging”. In: *NeuroImage* 108 (2015), pp. 441–449. DOI: 10.1016/j.neuroimage.2014.12.080.

[80] KK Osen et al. “Waxholm Space atlas of the rat brain auditory system: Three-dimensional delineations based on structural and diffusion tensor magnetic resonance imaging”. In: *NeuroImage* 199 (2019), pp. 38–56. DOI: 10.1016/J.NEUROIMAGE.2019.05.016.

[81] JFP Ullmann et al. “An MRI atlas of the mouse basal ganglia”. In: *Brain Structure and Function* 219.4 (2014), pp. 1343–1353. DOI: 10.1007/s00429-013-0572-0.

[82] C Watson et al. “An ontologically consistent MRI-based atlas of the mouse diencephalon”. In: *NeuroImage* 157 (2017), pp. 275–287. DOI: 10.1016/j.neuroimage.2017.05.057.

[83] Q Wang et al. “The Allen Mouse Brain Common Coordinate Framework: A 3D reference atlas”. In: *Cell* 181 (2020), pp. 1–18. DOI: 10.1016/j.cell.2020.04.007.

[84] TC Murakami et al. “A three-dimensional single-cell-resolution whole-brain atlas using CUBIC-X expansion microscopy and tissue clearing”. In: *Nature Neuroscience* 21.4 (2018). DOI: 10.1038/s41593-018-0109-1.

[85] Y Kim et al. “Brain-wide maps reveal stereotyped cell-type-based cortical architecture and subcortical sexual dimorphism”. In: *Cell* 171.2 (2017), pp. 456–469. DOI: 10.1016/j.cell.2017.09.020.

[86] SC Yates and M Puchades. *Extraction of parvalbumin positive cells from an Allen mouse brain in situ hybridisation experiment*. Human Brain Project Neuroinformatics Platform, 2019. DOI: 10.25493/6DYS-M3W.

[87] G Paxinos and C Watson. *The rat brain in stereotaxic coordinates*. Academic Press, 1982.

[88] HJJM Van De Werd and HBM Uylings. "Comparison of (stereotactic) parcellations in mouse prefrontal cortex". In: *Brain Structure & Function* 219.2 (2014), pp. 433–459. DOI: 10.1007/s00429-013-0630-7.

[89] M Bota and LW Swanson. "Collating and curating neuroanatomical nomenclatures: principles and use of the Brain Architecture Knowledge Management System (BAMS)". In: *Frontiers in Neuroinformatics* 4.3 (2010), pp. 1–16. DOI: 10.3389/fninf.2010.00003.

[90] M Laubach et al. "What, if anything, is rodent prefrontal cortex?" In: *eNeuro* 5.5 (2018), pp. 1–14. DOI: 10.1523/ENEURO.0315-18.2018.

[91] JK Mai and M Majtanik. "Toward a common terminology for the Thalamus". In: *Frontiers in Neuroanatomy* 12.114 (2019), pp. 1–23. DOI: 10.3389/fnana.2018.00114.

[92] JF Nylén and M Dorst. *Morphological reconstruction of cholinergic interneurons in the striatum*. Human Brain Project Neuroinformatics Platform, 2019. DOI: 10.25493/3EV4-TDG.

[93] C Boccara et al. *Sagittal and horizontal section images showing neuronal nuclei, calbindin and parvalbumin staining in the rat hippocampal region*. Human Brain Project Neuroinformatics Platform, 2019. DOI: 10.25493/JQ8F-TNF.

[94] N Berggaard et al. "Development of parvalbumin-expressing basket terminals in layer II of the rat medial entorhinal cortex". In: *eNeuro* 5.3 (2018). DOI: 10.1523/ENEURO.0438-17.2018.

[95] IE Bjerke et al. *Database of quantitative cellular and subcellular morphological properties from rat and mouse basal ganglia*. Human Brain Project Neuroinformatics Platform, 2019. DOI: 10.25493/2FDYXZ-76U.

[96] G Paxinos et al. *MRI/DTI Atlas of the Rat Brain*. 1st ed. Academic Press Inc, 2015.

[97] IE Bjerke et al. *Comparability of basal ganglia delineations across different rat brain atlases*. Human Brain Project Neuroinformatics Platform, 2019. DOI: 10.25493/D2M9-BSK.

[98] IE Bjerke et al. *Comparability of basal ganglia delineations across different mouse brain atlases*. Human Brain Project Neuroinformatics Platform, 2019. DOI: 10.25493/MWAS-3S6.

[99] IE Bjerke et al. *Franklin & Paxinos' "The Mouse Brain in Stereotaxic Coordinates" (3rd edition) spatially registered to the Allen mouse brain Common Coordinate Framework*. Human Brain Project Neuroinformatics Platform, 2019. DOI: 10.25493/1BT9-YYD.

[100] IE Bjerke et al. *Franklin & Paxinos' "The Mouse Brain in Stereotaxic Coordinates" (4th edition) spatially registered to the Allen Mouse Common Coordinate Framework*. Human Brain Project Neuroinformatics Platform, 2019. DOI: 10.25493/WFCZ-FSN.

[101] IE Bjerke et al. *Paxinos & Franklin's "The Mouse Brain in Stereotaxic Coordinates" (2nd edition) spatially registered to the Allen Mouse Common Coordinate Framework*. Human Brain Project Neuroinformatics Platform, 2019. DOI: 10.25493/BTKK-CRY.

[102] IE Bjerke et al. *Paxinos & Watson's "The Rat Brain in Stereotaxic Coordinates" (3rd edition) spatially registered to the Waxholm Space atlas of the rat brain*. Human Brain Project Neuroinformatics Platform, 2019. DOI: 10.25493/KNB2-GMN.

[103] IE Bjerke et al. *Paxinos & Watson's "The Rat Brain in Stereotaxic Coordinates" (4th edition) spatially registered to the Waxholm Space atlas of the rat brain*. Human Brain Project Neuroinformatics Platform, 2019. DOI: 10.25493/W3R1-R4A.

[104] IE Bjerke et al. *Paxinos & Watson's "The Rat Brain in Stereotaxic Coordinates" (5th edition) spatially registered to the Waxholm Space atlas of the rat brain*. Human Brain Project Neuroinformatics Platform, 2019. DOI: 10.25493/KQ5K-S0D.

[105] IE Bjerke et al. *Paxinos & Watson's "The Rat Brain in Stereotaxic Coordinates" (6th edition) spatially registered to the Waxholm Space atlas of the rat brain*. Human Brain Project Neuroinformatics Platform, 2019. DOI: 10.25493/XQ8J-TNE.

[106] IE Bjerke et al. *Paxinos & Watson's "The Rat Brain in Stereotaxic Coordinates" (7th edition) spatially registered to the Waxholm Space atlas of the rat brain*. Human Brain Project Neuroinformatics Platform, 2019. DOI: 10.25493/APWV-37H.

[107] IE Bjerke et al. *Swanson's "Brain Maps: Structure of the Rat Brain" (1st edition) spatially registered to the Waxholm Space atlas of the rat brain*. Human Brain Project Neuroinformatics Platform, 2019. DOI: 10.25493/ZB03-H5G.

[108] IE Bjerke et al. *Swanson's "Brain Maps: Structure of the Rat Brain" (2nd edition) spatially registered to the Waxholm Space atlas of the rat brain*. Human Brain Project Neuroinformatics Platform, 2019. DOI: 10.25493/EEQA-9RM.

[109] IE Bjerke et al. *Swanson's "Brain Maps: Structure of the Rat Brain" (3rd edition) spatially registered to the Waxholm Space atlas of the rat brain*. Human Brain Project Neuroinformatics Platform, 2019. DOI: 10.25493/ZFXB-23F.

[110] IE Bjerke et al. *Swanson's "Brain Maps: Structure of the Rat Brain" (4th edition) spatially registered to the Waxholm Space atlas of the rat brain*. Human Brain Project Neuroinformatics Platform, 2019. DOI: 10.25493/486N-966.

[111] A Laja et al. *Distribution of parvalbumin-positive interneurons in the normal adult mouse brain*. EBRAINS, 2020. DOI: 10.25493/BXGX-WM4.

[112] A Laja et al. *Distribution of parvalbumin-positive interneurons in the normal adult rat brain*. EBRAINS, 2020. DOI: 10.25493/8KCQ-3C7.

[113] IE Bjerke et al. *Brain-wide quantitative data on parvalbumin positive neurons in the rat*. EBRAINS, 2020. DOI: 10.25493/KR92-C33.

[114] IE Bjerke et al. *Brain-wide quantitative data on parvalbumin positive neurons in the mouse*. EBRAINS, 2020. DOI: 10.25493/BT8X-FN9.

[115] IE Bjerke and TB Leergaard. *Distribution of calbindin positive neurons in the normal adult mouse brain*. EBRAINS, 2020. DOI: 10.25493/KHNT-KV8.

[116] IE Bjerke et al. *Brain-wide quantitative data on calbindin positive neurons in the mouse*. EBRAINS, 2020. DOI: 10.25493/TT2Y-23N.

[117] S Lillehaug et al. "Data descriptor: Brain-wide distribution of reporter expression in five transgenic tetracycline-transactivator mouse lines". In: *Scientific Data* 6 (2019). DOI: 10.1038/sdata.2019.28.

[118] XJ Chen et al. "Neuroanatomical differences between mouse strains as shown by high-resolution 3D MRI". In: *NeuroImage* 29.1 (2006), pp. 99–105. DOI: 10.1016/j.neuroimage.2005.07.008.

[119] J Scholz et al. "Variability of brain anatomy for three common mouse strains". In: *NeuroImage* 142 (2016), pp. 656–662. DOI: 10.1016/j.neuroimage.2016.03.069.

[120] AM Khan et al. "Computer vision evidence supporting craniometric alignment of rat brain atlases to streamline expert-guided, first-order migration of hypothalamic spatial datasets related to behavioral control". In: *Frontiers in Systems Neuroscience* 12.7 (2018), pp. 1–29. DOI: 10.3389/fnsys.2018.00007.

[121] DE Oorschot. "Total number of neurons in the neostriatal, pallidal, subthalamic, and substantia nigral nuclei of the rat basal ganglia: A stereological study using the cavalieri and optical disector methods". In: *Journal of Comparative Neurology* 366 (1996), pp. 580–599. DOI: 10.1002/(SICI)1096-9861(19960318)366:4<580::AID-CNE3>3.0.CO;2-0.

[122] ZC Baquet et al. "A comparison of model-based (2D) and design-based (3D) stereological methods for estimating cell number in the substantia nigra pars compacta (SNpc) of the C57BL/6J mouse". In: *Neuroscience* 161.4 (2009), pp. 1082–1090. DOI: 10.1016/j.neuroscience.2009.04.031.

[123] Kristin Hamre et al. "Differential strain susceptibility following 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) administration acts in an autosomal dominant fashion: quantitative analysis in seven strains of Mus musculus". In: *Brain Research* 828.1-2 (1999), pp. 91–103. DOI: 10.1016/S0006-8993(99)01273-1.

[124] M Abercrombie. "Estimation of nuclear population from microtome sections". In: *The Anatomical Record* 94.2 (1946), pp. 239–247. DOI: 10.1002/ar.1090940210.

[125] C Schmitz and PR Hof. "Design-based stereology in neuroscience". In: *Neuroscience* 130.4 (2005), pp. 813–831. DOI: 10.1016/j.neuroscience.2004.08.050.

[126] MJ West. "Stereological methods for estimating the total number of neurons and synapses: Issues of precision and bias". In: *Trends in Neurosciences* 22.2 (1999), pp. 51–61. DOI: 10.1016/S0166-2236(98)01362-9.

[127] PGH Clarke. "How inaccurate is the Abercrombie correction factor for cell counts?" In: *Trends in Neurosciences* 15.6 (1992), pp. 211–212. DOI: 10.1016/0166-2236(92)90036-8.

[128] CS Von Bartheld. "Comparison of 2-D and 3-D counting: The need for calibration and common sense". In: *Trends in Neurosciences* 24.9 (2001), pp. 504–506. DOI: 10.1016/S0166-2236(00)01960-3.

[129] RW Guillery. "On counting and counting errors". In: *Journal of Comparative Neurology* 447.1 (2002), pp. 1–7. DOI: 10.1002/cne.10221.

[130] E Meijering. "Cell segmentation: 50 Years down the road". In: *IEEE Signal Processing Magazine* 29.5 (2012), pp. 140–145. DOI: 10.1109/MSP.2012.2204190.

[131] A Kan. "Machine learning applications in cell image analysis". In: *Immunology and Cell Biology* 95.6 (2017), pp. 525–530. DOI: 10.1038/icb.2017.16.

[132] L Silvestri et al. "Quantitative neuroanatomy of all Purkinje cells with light sheet microscopy and high-throughput image analysis." In: *Frontiers in Neuroanatomy* 9.68 (2015), pp. 1–11. DOI: 10.3389/fnana.2015.00068.

[133] SM Attili et al. "Cell numbers, distribution, shape, and regional variation throughout the murine hippocampal formation from the adult brain Allen Reference Atlas". In: *Brain Structure and Function* 224 (2019), pp. 2883–2897. DOI: 10.1101/635201.

[134] N Pallast et al. "Atlas-based imaging data analysis tool for quantitative mouse brain histology (AIDAhisto)". In: *Journal of Neuroscience Methods* 326 (2019), p. 108394. DOI: 10.1016/j.jneumeth.2019.108394.

[135] ME Vandenberghe et al. "High-throughput 3D whole-brain quantitative histopathology in rodents". In: *Scientific Reports* 6.1 (2016), pp. 1–12. DOI: 10.1038/srep20958.

[136] D Fürth et al. "An interactive framework for whole-brain maps at cellular resolution". In: *Nature Neuroscience* 21.1 (2018), pp. 139–153. DOI: 10.1038/s41593-017-0027-7.

[137] SC Yates et al. "QUINT: Workflow for Quantification and Spatial Analysis of Features in Histological Images From Rodent Brain". In: *Frontiers in Neuroinformatics* 13.75 (2019), pp. 1–14. DOI: 10.3389/fninf.2019.00075.

[138] HW Dong. *Allen Reference Atlas: A digital color brain atlas of the C57BL/6J male mouse*. Hoboken, NJ: John Wiley & Sons, 2008.

[139] WJ Bug et al. "The NIFSTD and BIRNLex vocabularies: building comprehensive ontologies for neuroscience." In: *Neuroinformatics* 6.3 (2008), pp. 175–194. DOI: 10.1007/s12021-008-9032-z.

[140] L French and P Pavlidis. "Using text mining to link journal articles to neuroanatomical databases." In: *The Journal of comparative neurology* 520.8 (2012), pp. 1772–1783. DOI: 10.1002/cne.23012.

[141] J Menke et al. "Rigor and Transparency Index, a new metric of quality for assessing biological and medical science methods". In: *bioRxiv* (2020). DOI: 10.1101/2020.01.15.908111.

[142] I Zaslavsky, RA Baldock, and J Boline. "Cyberinfrastructure for the digital brain: spatial standards for integrating rodent brain atlases." In: *Frontiers in Neuroinformatics* 8.74 (2014). DOI: 10.3389/fninf.2014.00074.

[143] GA Johnson et al. "Waxholm Space: An image-based reference for coordinating mouse brain research". In: *NeuroImage* 53.2 (2010), pp. 365–372. DOI: 10.1016/J.NEUROIMAGE.2010.06.067.

[144] GA Ascoli. "Mobilizing the base of neuroscience data: The case of neuronal morphologies". In: *Nature Reviews Neuroscience* 7.4 (2006), pp. 318–324. DOI: 10.1038/nrn1885.

[145] H Akil, ME Martone, and DC Van Essen. "Challenges and opportunities in mining neuroscience data". In: *Science* 331.6018 (2011), pp. 708–12. DOI: 10.1126/science.1199305.

[146] RE Coggeshall. "A consideration of neural countings methods". In: *Trends in Neurosciences* 15.1 (1992), pp. 9–13. DOI: 10.1016/0166-2236(92)90339-A.

[147] DE Oorschot. "Are you using neuronal densities, synaptic densities or neurochemical densities as your definitive data? There is a better way to go". In: *Progress in Neurobiology* 44.3 (1994), pp. 233–247. DOI: 10.1016/0301-0082(94)90040-X.

[148] R Egger et al. "Generation of dense statistical connectomes from sparse morphological data." In: *Frontiers in Neuroanatomy* 8.129 (2014), pp. 1–18. DOI: 10.3389/fnana.2014.00129.

[149] Marianne J Bezaire and Ivan Soltesz. "Quantitative assessment of CA1 local circuits: knowledge base for interneuron-pyramidal cell connectivity." In: *Hippocampus* 23.9 (2013), pp. 751–785. DOI: 10.1002/hipo.22141.

[150] P Tiesinga et al. "Feeding the human brain model". In: *Current Opinion in Neurobiology* 32 (2015), pp. 107–114. DOI: 10.1016/j.conb.2015.02.003.

[151] CF Taylor et al. "The minimum information about a proteomics experiment (MIAPE)". In: *Nature Biotechnology* 25.8 (2007), pp. 887–893. DOI: 10.1038/nbt1329.

[152] M Jessop, M Weeks, and J Austin. "CARMEN: A practical approach to metadata management". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 368.1926 (2010), pp. 4147–4159. DOI: 10.1098/rsta.2010.0147.

[153] AE Bandrowski and ME Martone. "RRIDs: A simple step toward improving reproducibility through rigor and transparency of experimental methods". In: *Neuron* 90 (2016), pp. 434–436. DOI: 10.1016/j.neuron.2016.04.030.

[154] E Marcus. "A STAR Is Born". In: *Cell* 166.5 (2016), pp. 1059–1060. DOI: 10.1016/j.cell.2016.08.021.

[155] A Keshavan and JB Poline. "From the wet lab to the web lab: A paradigm shift in brain imaging research". In: *Frontiers in Neuroinformatics* 13.3 (2019), pp. 1–13. DOI: 10.3389/fninf.2019.00003.

[156] KE Bouchard et al. "High-performance computing in neuroscience for data-driven discovery, integration, and dissemination". In: *Neuron* 92 (2016), pp. 628–631. DOI: 10.1016/j.neuron.2016.10.035.

[157] GA Ascoli et al. “Win-win data sharing in neuroscience”. In: *Nature Methods* 14.2 (2017), pp. 112–116. DOI: 10.1038/nmeth.4152.

[158] AR Ferguson et al. “Big data from small data: Data-sharing in the ’long tail’ of neuroscience”. In: *Nature Neuroscience* 17.11 (2014), pp. 1442–1447. DOI: 10.1038/nn.3838.

[159] F Leitner et al. “Data publications correlate with citation impact”. In: *Frontiers in Neuroscience* 10.419 (2016), pp. 1–6. DOI: 10.3389/fnins.2016.00419.

[160] T Bloom, E Ganley, and M Winker. “Data Access for the Open Access Literature: PLOS’s Data Policy”. In: *PLoS Medicine* 11.2 (2014), e1001607. DOI: 10.1371/journal.pmed.1001607.

[161] LM Federer et al. “Data sharing in PLOS ONE: An analysis of data availability statements”. In: *PLoS ONE* 13.5 (2018). DOI: 10.1371/journal.pone.0194768.

[162] G Colavizza et al. “The citation advantage of linking publications to research data”. In: *PLOS ONE* 15.4 (2020). DOI: 10.1371/journal.pone.0230416.

[163] BE Bierer, M Crosas, and HH Pierce. “Data Authorship as an Incentive to Data Sharing”. In: *New England Journal of Medicine* 376.17 (2017), pp. 1684–1687. DOI: 10.1056/NEJMsb1616595.

[164] HH Pierce et al. “Credit data generators for data reuse”. In: *Nature* 570.7759 (2019), pp. 30–32. DOI: 10.1038/d41586-019-01715-4.

[165] A Kreshuk and C Zhang. “Machine Learning: Advanced Image Segmentation Using ilastik”. In: *Computer Optimized Microscopy: Methods and Protocols*. Ed. by E Rebollo and M Bosch. Humana, New York, NY, 2019. Chap. 21, pp. 449–463. DOI: 10.1007/978-1-4939-9686-5{_}21.

Papers

I

frontiers in Neuroanatomy

ORIGINAL RESEARCH
published: 02 November 2018
doi: 10.3389/fnana.2018.00082

Navigating the Murine Brain: Toward Best Practices for Determining and Documenting Neuroanatomical Locations in Experimental Studies

*Ingvild E. Bjerke, Martin Øvsthus, Krister A. Andersson, Camilla H. Blixhavn, Heidi Kleven, Sharon C. Yates, Maja A. Puchades, Jan G. Bjaalie and Trygve B. Leergaard**

Department of Molecular Medicine, Institute of Basic Medical Sciences, University of Oslo, Oslo, Norway

In experimental neuroscientific research, anatomical location is a key attribute of experimental observations and critical for interpretation of results, replication of findings, and comparison of data across studies. With steadily rising numbers of publications reporting basic experimental results, there is an increasing need for integration and synthesis of data. Since comparison of data relies on consistently defined anatomical locations, it is a major concern that practices and precision in the reporting of location of observations from different types of experimental studies seem to vary considerably. To elucidate and possibly meet this challenge, we have evaluated and compared current practices for interpreting and documenting the anatomical location of measurements acquired from murine brains with different experimental methods. Our observations show substantial differences in approach, interpretation and reproducibility of anatomical locations among reports of different categories of experimental research, and strongly indicate that ambiguous reports of anatomical location can be attributed to missing descriptions. Based on these findings, we suggest a set of minimum requirements for documentation of anatomical location in experimental murine brain research. We furthermore demonstrate how these requirements have been applied in the EU Human Brain Project to optimize workflows for integration of heterogeneous data in common reference atlases. We propose broad adoption of some straightforward steps for improving the precision of location metadata and thereby facilitating interpretation, reuse and integration of data.

Keywords: best practice, brain atlas, data mining, data sharing, FAIR, reproducibility, location metadata, rodent brain

OPEN ACCESS

Edited by:
Hans J. ten Donkelaar, Radboud University Nijmegen, Netherlands

Reviewed by:
Piotr Majka, Nencki Institute of Experimental Biology (PAS), Poland
Robert P. Vertes, Florida Atlantic University, United States

****Correspondence:***
Trygve B. Leergaard
t.b.leergaard@medisin.uio.no

Received: *13 July 2018*
Accepted: *19 September 2018*
Published: *02 November 2018*

Citation:
Bjerke IE, Øvsthus M, Andersson KA, Blixhavn CH, Kleven H, Yates SC, Puchades MA, Bjaalie JG and Leergaard TB (2018) Navigating the Murine Brain: Toward Best Practices for Determining and Documenting Neuroanatomical Locations in Experimental Studies. Front. Neuroanat. 12:82. doi: 10.3389/fnana.2018.00082

INTRODUCTION

Over last decades, considerable effort has been invested in large-scale the production of neuroscience data (see, e.g., Stopps et al., 2004; Boy et al., 2006; Lein et al., 2007; Zakiewicz et al., 2011; Hintiryan et al., 2012; Ragan et al., 2012; Oh et al., 2014). With increasingly efficient data production pipelines the number of scientific reports and amount of available data is steadily growing (Hey and Trefethen, 2003). To organize, compare and integrate such large amounts of

data into new knowledge and understanding about the brain, new computational approaches have emerged (Amari et al., 2002; Bjaalie et al., 2005; Koslow and Subramaniam, 2005; Bjaalie, 2008; Tiesinga et al., 2015; Bjerke et al., 2018) to make data discoverable, accessible, interpretable and re-usable, as outlined in the widely endorsed FAIR Guiding Principles (Findability, Accessibility, Interoperability, and Re-usability; Wilkinson et al., 2016). However, these integration efforts face the challenge that neuroscience data span multiple spatial and temporal scales (see, e.g., Amunts et al., 2016), and that results are commonly reported in journal articles as narratives supported with documentation in selected figures and tables that are difficult to compare. A prerequisite for data integration is that the nature and relationships of data parameters are well defined and easily comparable, hence integration efforts will have to incorporate methods for dealing with these differences.

Interpretation of observations collected from the brain depends critically on specific information about anatomical location (see, e.g., Bjaalie, 2002): e.g., from which cortical area, cell layer, or nucleus were measurements or observations obtained? Comparison of results across studies, or replication of experimental findings, necessitates that the specific anatomical position of a measurement, observation, or experimental perturbation, is well-defined. Such anatomical descriptions in experimental reports are of variable quality and are prone to ambiguity, since anatomical terms can be interpreted in a number of ways, and alternative anatomical parcellation schemes often uses different boundary definitions (see, e.g., Van De Werd and Uylings, 2014). Thus, the lack of universally accepted and well-defined descriptions of neuroanatomical location, defining the precise location being studied, is a major challenge when attempting to compare and integrate data from different investigations.

In response to this challenge, open access, three-dimensional (3-D) brain atlases have been developed for murine brains (Hjornevik et al., 2007; Lein et al., 2007; Johnson et al., 2010; Hawrylycz et al., 2011; Veraart et al., 2011; Oh et al., 2014; Papp et al., 2014; Kjonigsen et al., 2015) to serve as spatial frameworks for data sharing and integration (Boline et al., 2008; Amunts et al., 2014; Zaslavsky et al., 2014). Building upon a new generation of 3-D brain atlases, the EU Human Brain Project develops tools and workflows for integrating, sharing and analyzing brain data that have been defined within a common anatomical framework. The project has established workflows for mapping diverse types of murine and human image data to common spatial reference atlas frameworks, building on tools for spatial registration of two-dimensional (2-D) histological image data to a 3-D reference volume (Papp et al., 2016; Puchades et al., 2017), and use of organized collections of metadata describing basic features of data, including descriptions of the anatomical location from which the data originate. These tools and workflows are currently routinely used in the Human Brain Project, but their wider adoption outside the project requires a better understanding of the presently used approaches for describing and documenting neuroanatomical location in experimental studies of murine brains.

To first assess how anatomical location is reported in the neuroscience literature, we evaluated and compared current practices for interpreting and documenting the location of measurements in different disciplines of neuroscience that typically deal with invasive techniques or extraction of tissue. Our observations indicate substantial differences in approach, interpretation and reproducibility of anatomical location between different categories of experimental research, as well as a potential for improvement with relatively simple measures. Based on these observations, we have adjusted and optimized the Human Brain Project tools and workflows to accommodate the type of data and documentation typically used in different domains of experimental research on murine brains. We propose step-wise practical implementations that can improve current practices, and argue that these procedures increase reproducibility and facilitate integration of neuroscience data. We finally discuss costs and benefits of increasingly elaborate approaches.

MATERIALS AND METHODS

Survey of Current Practices for Assignation and Reporting of Anatomical Location

We performed a literature study to explore current practices for reporting anatomical information in different categories of experimental neuroscientific studies in murine brains.

We focused on experimental studies involving invasive procedures or tissue extraction, and classified publications into the following seven methodological categories based on the principal methodology employed: (1) cytoarchitectonic staining techniques, including immunohistochemistry (IHC) and *in situ* hybridization (ISH); (2) axonal tract tracing; (3) transmission electron microscopy (TEM); (4) immunoblotting; (5) *in vitro* electrophysiology with slice preparations and microscopic visualization of recorded cells; (6) *in vivo* electrophysiology, and (7) two-photon and optogenetic imaging. Studies involving tomographic whole-brain imaging and trans-cranial measurements were not included.

Individual search strings were made for each methodological category, and a search was performed in Ovid MEDLINE. Except for the terms related to the specific methods, the criteria used for building the search strings were consistent across searches and contained the following: ((exp mice/or exp rats/) OR (mouse or mice or rat or rats).tw,kf.) AND ((brain or brains or neuroscien* or neuroanatom* or neuro anatom* or neuron or neurons).mp.). Strings related to the specific methodologies of interest (see above) were added to this:

(1) (immunohistochemistry/or immunohistochemistry.tw,kf.) OR ((in situ hybridization).tw,kf.)
(2) ((retrograde or anterograde) adj trac*.tw,kf.) for axonal tract tracing
(3) ((Microscopy, Electron, Transmission/) OR (transmission adj (electronmicroscop* or electron microscop*)).tw,kf.)
(4) ((western blot* or immunoblot*).tw,kf.)

(5) ((((invitro or in vitro) adj2 (electrophysiolog* or electro physiolog* or cell recording)) or cell recording).tw,kf.)
(6) ((invivo or in vivo) adj2 (electrophysiolog* or electro physiolog*)).tw,kf
(7) ((optogenetics/or optogenetics or optogenetic*.tw,kf.)) OR ((((twophoton or two photon or two-photon or 2 photon or 2-photon) adj2 (microscop* or imaging)).tw,kf.))

Filters were then added to limit results to those with journal article format and publication data from 2012 through 15.02.2017. The search returned 9839 entries related to immunohistochemistry and *in situ* hybridization, 547 related to axonal tract tracing, 949 related to electron microscopy, 7004 related to immunoblotting, 95 related to *in vitro* electrophysiology, 213 related to *in vivo* electrophysiology, and 2023 related to optogenetic or two-photon microscopy.

Papers (n = 120; 20 for each methodological category) were chosen from the selection of search entries by use of a random number generator and evaluated using the following inclusion criteria: (a) contained murine brain data; (b) presented original data; and (c) were published within the last 5 years. Papers not meeting these criteria were excluded and a new random paper selected.

For each paper in the survey, we evaluated the descriptions of anatomical locations with respect to: (1) any additional information provided beyond the structure name (e.g., by citing an anatomical reference atlas, illustration of the region of interest by use of a schematic drawing or reference atlas plate, or description of the general histological, cytoarchitectonic or electrophysiological features of the region); (2) use of histological sections (without counterstaining); (3) use of (immuno-)histochemical staining to visualize anatomical features; (4) specification of spatial coordinates (e.g., stereotaxic coordinates observed during experimental surgery or by comparison with a reference atlas); (5) documentation using images that show anatomical landmarks suitable for identifying location in addition to features of interest (see below); (6) annotation of anatomical landmarks or boundaries in images from the material; (7) images from multiple (serial) sections through a region of interest; and (8) spatial registration of images to a reference atlas.

Some papers reported results obtained using several methodologies, but for each paper we only assessed the documentation related to the specific methodology for which the paper was selected. Documentation of anatomical location with images was only considered sufficient if images gave a reasonable overview of the regions of interest, allowing the reader to identify the position of the image relative to a reference atlas. We set the minimum standard to be that images should show the region of interest *and* at least one other distinct anatomical landmark, such as a part of the ventricular system, a major white matter tract, or a distinct gray matter structure. Consequently, high-power images showing structural details of a smaller region, e.g., a part of the cerebral cortex with visible layers, were not considered sufficient to allow interpretation of anatomical location in this context.

Most commonly, the region of interest was an *observation site* in which some analysis had been performed (e.g., cell counting, immunoreactivity observations, cell reconstructions). Alternatively, the region of interest may have been a site of an experimental procedure, or *perturbation* (e.g., a lesion, an electrode implantation or a virus injection). In the case of multiple regions of interest of the same type, e.g., cell counting in several regions or multi-site electrode recordings, we assumed the same level of effort had been undertaken to determine the location of each site, and we evaluated the paper according to the best documented region. In the case of multiple regions of interest of different types, we evaluated both types of regions separately. In tract tracing studies, for example, there are sites of perturbation (injection of tracer) and observation (labeled features). For tracing studies we therefore assessed injection sites and terminal fields as individual reports of regions of interest. Thus, while 20 papers from each of the seven methodological categories were surveyed, the total number of papers used was 120, because each tract tracing paper was included in two of the categories (tract tracing injection site and tract tracing terminal fields). References to these 120 papers are given in **Supplementary Table 1**, which also provides an overview of the observations extracted from 162 different reports of anatomical observations found in these papers.

Tools Used to Facilitate Documentation of Anatomical Locations

We aimed to demonstrate how new tools and procedures can be applied in order to map and co-visualize data spanning several methodological categories, and to identify key strategies in this process that should influence how data are acquired and documented.

We used the Human Brain Project software tool QuickNII (Puchades et al., 2017) to register single or serial section images to a 3-D reference atlas template by positioning and slicing the atlas in user-defined planes of sectioning. The QuickNII tool is bundled either with the Waxholm Space atlas of the Sprague-Dawley rat brain (version 2, Papp et al., 2014; Kjonigsen et al., 2015)[1] or the Allen Mouse Common Coordinate Framework (version 3, downloaded June 17, 2016; Oh et al., 2014)[2]. We furthermore used the Human Brain Project tool LocaliZoom for extraction of coordinates from annotated points of interest. The coordinates representing the location of these points in reference atlas space were exported as x, y, z coordinates to MeshView, a Human Brain Project web-viewer tool for visualization of 3-D mesh-data (structural atlas parcellations) together with the point coordinates extracted with LocaliZoom.

Data Used to Demonstrate Workflows

To demonstrate workflows for spatial integration of different types of data, we used existing or publicly available data sets from the following four methodological categories: (1) *in vivo* electrophysiology; (2) transmission electron microscopy; (3) cytoarchitectonic staining techniques; and (4) *in vitro* electrophysiology with cell reconstruction. These categories

[1]https://www.nitrc.org/projects/whs-sd-atlas

[2]http://download.alleninstitute.org/informatics-archive/current-release/mouse_ccf/annotation/ccf_2015/

represent four of the seven methodologies included in our literature survey, and the workflows used to map these data to anatomical space can easily be extended to the remaining categories.

Electron microscopy data showing parvalbumin positive neurons in the rat medial entorhinal cortex (Berggaard et al., 2018), was generously made available to the present study by Nina Berggaard (Norwegian University of Science and Technology, Norway). *In vivo* electrophysiology recording data from the rat hippocampal region were produced by Debora Ledergerber (Norwegian Institute of Science and Technology, Norway; Puchades et al., 2017). Immunohistological material showing parvalbumin positive neurons across a horizontally cut hemisphere (Boccara et al., 2015) was shared through the Human Brain Project by Menno P. Witter (Norwegian University of Science and Technology, Norway). A 3-D reconstruction of a mouse striatal cholinergic interneuron was performed by Alexander Kozlov, Johanna Frost Nylén and Sten Grillner (Karolinska Institutet, Sweden) and shared via the Human Brain Project[3]. Lastly, a series of sagittal sections from the Allen Institute for Brain Science repository of *in situ* hybridization data (Lein et al., 2007)[4] was downloaded through their API.

For each data set, section images were spatially registered to a reference atlas template using the QuickNII tool. The first three data sets (*in vivo* electrophysiology, immunohistochemistry, and electron microscopy data) were registered to the Waxholm Space atlas of the rat brain. Section material from the *in vitro* electrophysiology with cell reconstruction and *in situ* hybridization was mapped to the Allen Mouse Common Coordinate Framework of the mouse brain. Following spatial registration to atlas, images with associated atlas information were exported to LocaliZoom for visualization and retrieval of spatial coordinates.

[3]https://www.humanbrainproject.eu/en/explore-the-brain/search/

[4]http://mouse.brain-map.org/experiment/show/75457579

Ethical Considerations

This study used animal data acquired in accordance with European Union and International legislation regarding use of animal subjects. For data shared by the Human Brain Project, verification of compliance with European legal and regulatory requirements is provided with the data. For other data, statements regarding ethical conduct care are found in the original papers (Lein et al., 2007; Berggaard et al., 2018).

RESULTS

Survey of Anatomical Descriptions and Metadata Provided in Original Neuroscientific Reports

We surveyed anatomical descriptions and documentation provided in 120 scientific original reports (published within the last 5 years) involving different types of experimental methods, and evaluated their inclusion of tissue sectioning and histological staining, specification of spatial coordinates, documentation with images (with or without annotations) and the use of spatial registration to anatomical reference atlases. We found systematic variations across methodological categories regarding the degree to which anatomical locations were described and documented (summarized in **Table 1**). Below, we first summarize our findings of anatomical documentation per methodological category, and secondly compare the use of anatomical descriptions and different types of documentation across the methodological categories.

Tract-tracing studies generally provide more anatomical documentation than studies using other methods (**Table 1**). In 85% of the papers investigated, the location of tracer injection

TABLE 1 | Overview of anatomical metadata elements provided in the publications investigated.

	Description of ROI	Sectioning	Staining	Coordinates	Image documentation	Annotations	>1 section image	Atlas registration
Tract tracing (injection site)	100	100	80	85	80	50	20	5
Tract tracing (terminal fields)	80	100	95	45	60	45	15	0
Cytoarchitectonic studies	45	95	95	30	60	30	10	0
In vivo electrophysiology	95	45	25	65	30	10	0	5
In vitro electrophysiology	70	100	30	15	20	15	0	0
Advanced imaging	86	64	32	50	18	14	0	0
Electron microscopy	30	25	10	10	5	0	0	0
Western blot	20	5	0	0	0	0	0	0
Average	66	67	46	38	34	21	6	1

Percentage of papers (n = 120), sorted by methodological category (rows), using different approaches (columns) to document anatomical location selected articles. Employed practices to identify and document anatomical location vary across the categories of investigations. Most studies describe location semantically, while the use of image documentation and spatial registration of images to reference atlases is limited.

sites was given by the perioperatively recorded stereotaxic coordinates used when placing the tracer injections, and 35% of these further documented injection sites with section coordinates upon verification of position. Injection sites were often documented with images showing anatomical location of the injection (80%), and in 60% of cases the regions of interest containing neuronal labeling were also documented with images (**Table 1**). Contrary to our expectation, we found that in most of the investigated papers reporting on tract tracing experiments (17 of 20 papers, 85%), the anatomical location of tracer injection sites was more thoroughly described and documented than the location of transported neuronal labeling in one or more remote brain regions (see **Table 1** for details).

While all reports from cytoarchitectonic investigations used histological techniques and almost all (18 of 20, 90%) presented images of microscopic observations, only 60% provided images with visible anatomical landmarks (**Table 1**), and very few (10%) included images showing their region(s) of interest across multiple sections. In several of the papers investigated we found it difficult to interpret and reproduce the investigated regions in a reference atlas.

Our results also show that *in vivo* electrophysiological experiments typically provide better documentation of anatomical location than most other study types, mainly as the location of the recording electrodes (in 65% of cases, **Table 1**) is usually defined by perioperatively recorded stereotaxic coordinates. In 38% of these papers, implantation sites were further documented by providing histologically verified section coordinates. Of the publications reporting on *in vitro* electrophysiological studies with microscopic visualization of recorded cells, 30% used histological staining to reveal anatomical landmarks, but only 20% included overview images documenting anatomical boundaries or landmarks (**Table 1**).

Studies using advanced *in vivo* optogenetic or two-photon microscopic imaging techniques often (in 85% of cases, **Table 1**) contained some form of description of the region of interest, and coordinates were provided in approximately half of the papers. Somewhat surprisingly, we found that the microscopic images of the analyzed material rarely (in 18% of the papers investigated, **Table 1**) included anatomical landmarks suitable for documenting anatomical locations.

Lastly, we found that documentation of anatomical location was, to a small degree, provided in reports of electron microscopy and immunoblot studies. 20–30% of such studies contained anatomical descriptions of regions of interest, while use of additional documentation was minimal or absent (**Table 1**). These types of studies were the least likely to include sufficient anatomical information of all the assessed methods.

Our results thus show that of all 120 papers surveyed only 66% included some form of anatomical descriptions of regions of interest, beyond mention of the region name (**Table 1**). In the remaining 34% of papers, we found no descriptions of the region of interest apart from the name of the region. A further breakdown of the 66% papers providing anatomical descriptions is summarized in **Table 2**. This breakdown showed that 29% of the papers providing a description of a region of interest did so by using a reference to a specific anatomical atlas. Notably, we also found that anatomical reference atlases were most frequently cited in reports of tract tracing injection sites (60%, **Table 2**), while none of the immunoblot reports providing anatomical descriptions related these to a reference atlas. Interestingly, among the anatomical descriptions provided in reports of neural labeling observed in tract-tracing studies, or advanced imaging studies, only 44 or 11%, respectively, included reference to a specific anatomical atlas. We further found that among the 66% of papers including anatomical descriptions, 76% included illustrations or line drawing of anatomical features, 42% indicated measurements of distances to specific anatomical landmarks, and 13% related their descriptions to observed microscopic or electrophysiological features (**Table 2**).

Although stereotaxic atlases are widely used and stereotaxic coordinates provide precise indications of location, we found

TABLE 2 | Overview of the types of descriptions and coordinate based information provided the publications investigated.

	Tract tracing injection site	Tract tracing terminal fields	Cytoarchitectonic studies	*In vivo* electrophysiology	*In vitro* electrophysiology	Advanced imaging	Electron microscopy	Western blot	Average
Descriptions	**100**	**80**	**45**	**95**	**70**	**86**	**30**	**20**	**66**
Based on distance to landmark	90	13	22	84	0	68	33	25	42
Based on reference atlas	60	44	22	53	21	11	17	0	29
Based on illustration	55	100	89	58	71	89	67	75	76
Based on cellular features	20	19	0	26	36	5	0	0	13
Coordinates	**85**	**45**	**30**	**65**	**15**	**50**	**10**	**0**	**38**
Point coordinates	100	0	0	100	0	91	50	0	43
Section coordinates	35	45	100	38	100	27	50	0	49

Percentage of papers (n = 120) from different methodological categories (columns) providing different types of documentation (rows) as descriptions or point coordinates.

that coordinate-based information was presented in only 38% of the papers surveyed (**Table 1**). Further breakdown showed that of the 38% of papers that included spatial coordinates, 49% provided positions of sections or slices reported as distances from skull landmarks or the midline (**Table 2**), while 43% specified points of interest as x, y, z coordinates (**Table 2**) targeted perioperatively and/or identified by *post hoc* analyses. Of the publications reporting tract tracing or *in vivo* electrophysiological experiments, 65% included spatial coordinates, while none of the publications reporting immunoblotting results contained such information.

While most studies included high-power image documentation of observed features, only 34% included images showing anatomical landmarks and/or boundaries suitable for interpretation of anatomical locations, while 21% of studies provided images with anatomical annotations superimposed (**Table 1**). Most of the images showing anatomical landmarks were restricted to one brain region of interest; in fact, only ∼9% of all studies provided overview images showing a whole, half or smaller part of a brain section. None of the papers examined included images from macroscopic dissection. Only 6% of the papers used more than a single section image to document the same region of interest (**Table 1**).

Only two of the surveyed papers (1%) used spatial registration tools to map the position of their experimental images to anatomical reference atlases (**Table 1**). We thus found systematic differences in the documentation of anatomical regions of interest provided in original research papers that varied across methodological subfields of neuroscience. Our findings indicate that most studies lack elementary descriptions and documentation of anatomical location that in principle should be straightforward to include in scientific reports, regardless of the type of methodology used.

Minimum Requirements for Documentation of Locations in Experimental Murine Brain Research

Based on the above findings, we considered how anatomical descriptions from different methodological traditions could be improved to achieve more consistent and reproducible descriptions of anatomical locations. A key principle underlying empirical scientific research is that original publications should contain sufficient descriptions of materials and methods used to allow peers to reproduce experimental results. Extrapolating from this, an obvious minimum requirement for the reporting of anatomical location is that anatomical regions of interest should be specifically and unambiguously reported, with sufficient documentation to allow interpretation and replication of described anatomical positions. However, our survey of the current literature above revealed that the location of data is often poorly described and documented, making reported anatomical positions hard to replicate. Combining the findings summarized in **Table 1** and accumulated experiences with interpretation and validation of anatomical locations in a wide range of materials measurements collected in context of the Human Brain Project (see e.g., Bjerke et al., 2018), we identified some key documentation elements that we found to be of particular importance for our ability to unequivocally specify anatomical locations for different data sets. We also formulate a set of method-independent recommendations for a minimum documentation practice that could alleviate the ambiguity observed in many research papers (see above), and facilitate interpretation of anatomical positions and comparison of research findings. Thus, to achieve more unambiguous and reproducible descriptions of anatomical locations in neuroscientific reports we propose that adherence to at least one and preferably several of the following recommendations should be set as a minimum requirement:

(1) Employ and refer to a specific anatomical parcellation scheme
(2) Provide precise semantic descriptions relating observations to anatomical landmarks or features
(3) Define points or regions of interest using an anatomical illustration or diagram
(4) Provide annotated images showing distinct anatomical landmarks
(5) Report spatial coordinates

Below, we specify and exemplify these recommendations in further detail.

Referring to a Defined Parcellation Scheme

Several parcellation schemes exist for the whole mouse and rat brain, including the widely used 2-D stereotaxic reference atlases (Swanson, 2004; Paxinos and Franklin, 2012; Paxinos and Watson, 2013), and 3-D reference atlas templates (Johnson et al., 2010; Hawrylycz et al., 2011; Papp et al., 2014). More detailed parcellation schemes have also been defined for parts of the brain, such as the hippocampus (Kjonigsen et al., 2011; Witter, 2012; Boccara et al., 2015). Anatomical parcellation schemes should preferably include graphical representations of the boundaries defining anatomical structures specified in a nomenclature list. Use of standardized schemes that are widely used in the community will facilitate comparison of anatomical locations. To be unambiguous, a description based on a parcellation scheme should include (1) the name of the region of interest exactly *as it appears* in the reference atlas, and (2) appropriate citation of the reference atlas (and version) employed. Reference to 3-D atlas templates, e.g., the Allen Mouse Common Coordinate Framework (Lein et al., 2007; Oh et al., 2014) or Waxholm Space (Hawrylycz et al., 2011; Papp et al., 2014), that can be sliced in any orientation provides superior anatomical precision for both volumetric data and 2-D sectioned data (**Figure 1**). For observations or measurements that are sampled from an entire brain region, for example describing populations of labeled cells distributed across a given brain region, reference to the region name will usually be unambiguous. However, if observations or measurements only pertain to a small subset of a region of interest, e.g., for a single cell reconstruction or a tissue sample processed for electron microscopy, information about parcellation scheme should be supplemented with one of the other recommendations listed above to more clearly specify

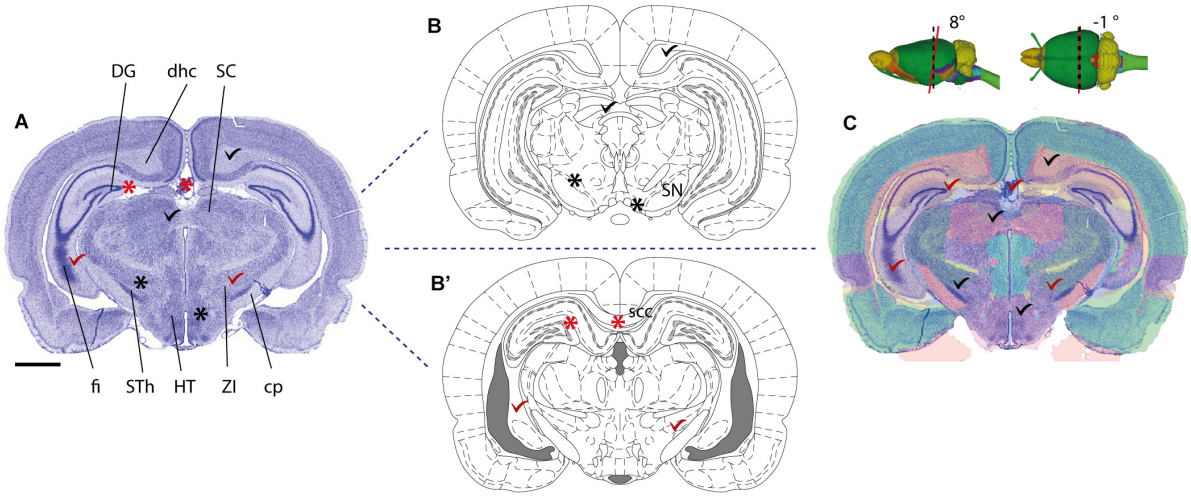


FIGURE 1 | Mapping a coronal rat brain section with a slight tilt to a reference atlas**. (A)** Image of a thionine stained coronal rat brain section, cut with a slight deviation from the standard plane of orientation. A slight left-right asymmetry is visible in the fimbria of the hippocampus (fi), and when comparing the section with a standard reference atlas, reference plates from different anteroposterior levels match the dorsal and ventral parts of the section, respectively. **(B,B')** Atlas diagrams 78 **(B)** and 68 **(B')**, reproduced from Paxinos and Watson (2005) with permission, separated by ~1.5 mm. In both diagrams, anatomical structures that are consistent (check marks) or inconsistent (asterisks) with corresponding regions in the histological section are seen. Regions located dorsally in **(A)**, such as the dentate gyrus (DG), dorsal hippocampal commissure (dhc) and the superior colliculus (SC) correspond with plate 78 **(B)**, but not plate 68 **(B')**, while regions located ventrally in **(A)**, such as the fimbria (fi), subthalamic nucleus (STh), hypothalamic region (HT), zona incerta (ZI), and cerebral peduncle (cp) correspond with plate 68 **(B')**, but not with plate 78 **(B)**. **(C)** A custom generated transparent atlas overlay from the 3-D Waxholm Space atlas of the rat brain (v2, Papp et al., 2014; Kjonigsen et al., 2015) superimposed onto the thionine section shows better overall anatomical correspondence of structures in both dorsal and ventral parts of the section. The location and tilt of the custom atlas plate is indicated by red lines in the inset 3-D figures. Scale bar, 2 mm.

location within the region of interest, e.g., using an image or anatomical illustration.

Semantically Describing Spatial Relation to Distinct Anatomical Landmarks or Architectonic Features

In some cases, a region of interest might be described by defining its relation to structural or cellular landmarks. This is a particularly relevant form of description when regions are defined differently or with more detail than in standard atlas frameworks. See, for example, Insausti et al. (1997, pp. 151–155), where subregional boundaries of the entorhinal cortex are described both in terms of cytoarchitectonic features, and in relation to anatomical landmarks. It should be emphasized that while such description can be elaborate and detailed, they can also be challenging to interpret without expert knowledge. Anatomical illustrations or annotated images can facilitate easier interpretation for the reader.

Indicating the Location in an Anatomical Illustration

Anatomical locations can be graphically defined using reference atlas diagrams, schematic summary drawings or other figures. Indication of sampling position within such illustrations (see, e.g., Akhter et al., 2014, their **Figure 2**) can serve as supplements to semantic descriptions, or alternative to spatial coordinates.

Providing (Annotated) Images

Images of experimental material may depict sections or macroscopic dissections. Low-resolution images are generally as useful for the purpose of visualizing location as high-resolution ones, even images obtained with a standard cell phone camera. Section images should show structural landmarks outside the region of interest, if possible more than one. Suggestions for anatomical landmarks that can be consistently identified in volumetric material of the rodent brain are provided in Sergejeva et al. (2015). For sectioned material, the size and shape of prominent gray and white matter regions (e.g., the hippocampus, caudoputamen, pontine nucleus, anterior commissure, and corpus callosum) are also highly useful in order to interpret location in the brain. Ideally, images should cover entire sections. A simple evaluation of the influence of the coverage of a section image on spatial registration accuracy (**Figure 2**) confirms that the more information an image contains, the more likely are two independent and equally experienced researchers to interpret the anatomical position of a region of interest consistently. For procedures not involving histological processing and tissue sections, macroscopic images of the whole brain or tissue sample(s) before and after dissection of tissue samples can improve the interpretation of the anatomical location of the investigated sample considerably (**Figure 3A**). Annotations defining regional boundaries and specifying locations sampled or measured increase precision considerably (see, e.g., Dobi et al., 2013; their **Figure 9**).

Using Spatial Coordinates

Spatial coordinates defined in relation to unique skull features or anatomical landmarks effectively communicate exact positions within the brain, independent of parcellation schemes. Descriptions based on spatial coordinates must specify the

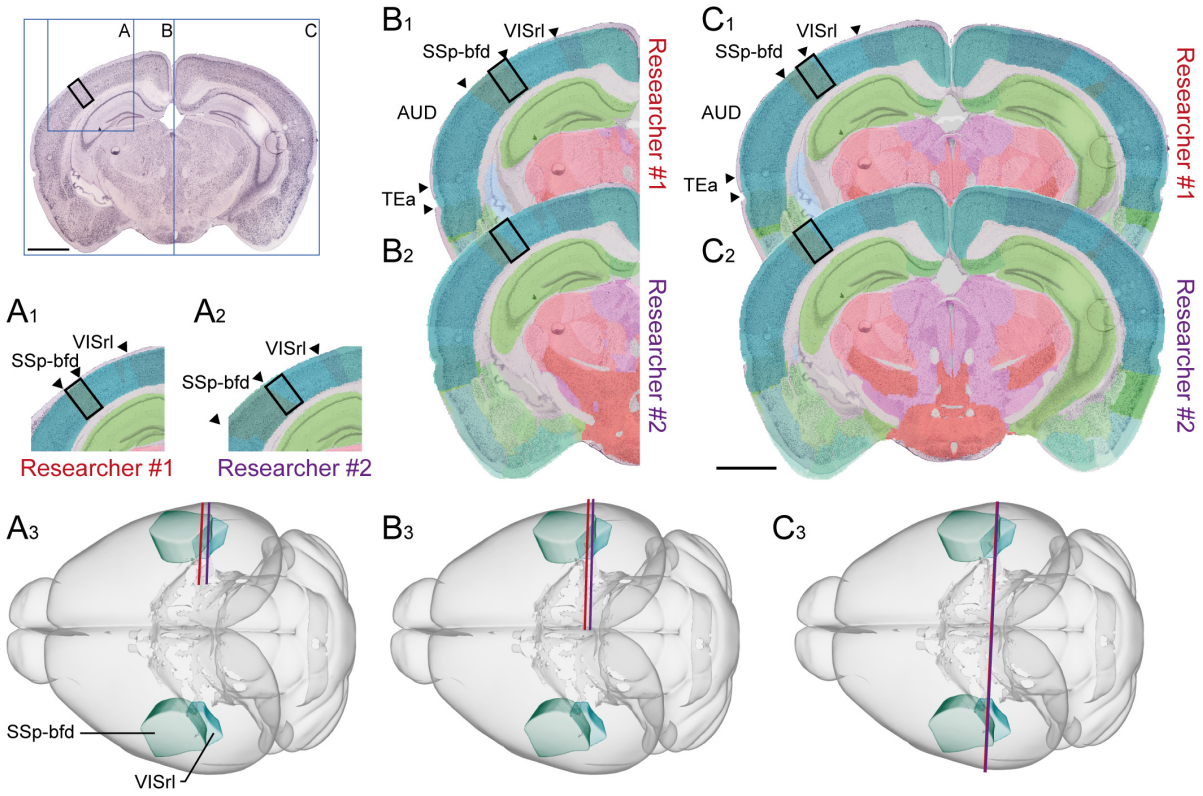


FIGURE 2 | Documenting anatomical location at varying levels of coverage. Illustration showing the influence of image coverage on the spatial accuracy of registration to a reference atlas. Upper left inset: image of a coronal, thionine stained mouse brain section, the black rectangle indicates an arbitrary region of interest in the cerebral cortex, blue rectangles indicate the size of the image used in examples **(A–C)**. **(A–C)** Images with increasing anatomical coverage were shown to two experienced researchers, who independently registered the images to the Allen Mouse Common Coordinate Framework using the QuickNII tool. **(A)** With only a small part of the section image available, the two researchers interpreted the location of the image differently, as the primary somatosensory cortex, barrel field (SSP-bfd, **A1**) by Researcher #1, and as the rostrolateral visual area (VISrl, **A2**) by the Researcher #2. (**A3**) Shows the positions of the section image assigned by Researcher #1 (red line) and Researcher #2 (purple line) in atlas space. **(B)** With an image showing half of the section, the researchers' interpretations of the position become more similar **(B1–B3)**, and when the entire section image was available, both researchers interpreted the region of interest to be located in SSp-bfd **(C1–C3)**. This illustrates that access to images covering more anatomical landmarks is important to provide reproducible information about anatomical positions. Scale bars: 2 mm.

reference space used (e.g., reference atlas or local coordinate system). Coordinates may indicate the level or distance of a section or slice from an anatomical landmark (e.g., bregma or the midline of the brain) or specific points defined by x, y, z coordinates. The method used to define coordinates should be specified and additional validation steps, such as e.g., histological confirmation of perioperative stereotaxic measurements supplemented with documentation using image(s) or illustrations, can improve precision.

Combining Different Types of Data in Reference Atlas Space

A considerable challenge for efforts toward integration of different types of neuroscience data is the heterogeneity in the spatial scale and modality of data. In context of the ambition of the Human Brain Project to make heterogeneous brain data accessible for integrative analyses and computational modeling (Bjerke et al., 2018), we have explored ways to assign location to disparate categories of murine neuroscience data. Using the minimum requirements for documentation of anatomical location proposed above as a starting point, and having the ambition to optimize anatomical descriptions of different types of neuroscience data, we established workflows to relate data sets acquired by *in vivo* electrophysiology, immunohistochemistry, *in situ* hybridization, transmission electron microscopy and *in vitro* electrophysiology with cell reconstruction to a common spatial atlas framework. The core workflow, used to spatially combine features of interest from different types of data sets, involves three steps. We first link the data to the same anatomical reference framework, secondly extract spatial coordinates representing features of interest from each of the data sets and thirdly co-visualize the extracted features in a 3-D atlas viewer as a starting point for various analytic approaches. The workflow is implemented using a suite of digital atlas and viewer tools developed in the Human Brain Project. The QuickNII tool is developed for registering 2-D (serial) images to a reference atlas by mapping a spatially corresponding, customized atlas image onto images (Puchades et al., 2017). The LocaliZoom viewer tool provides an overlay of custom

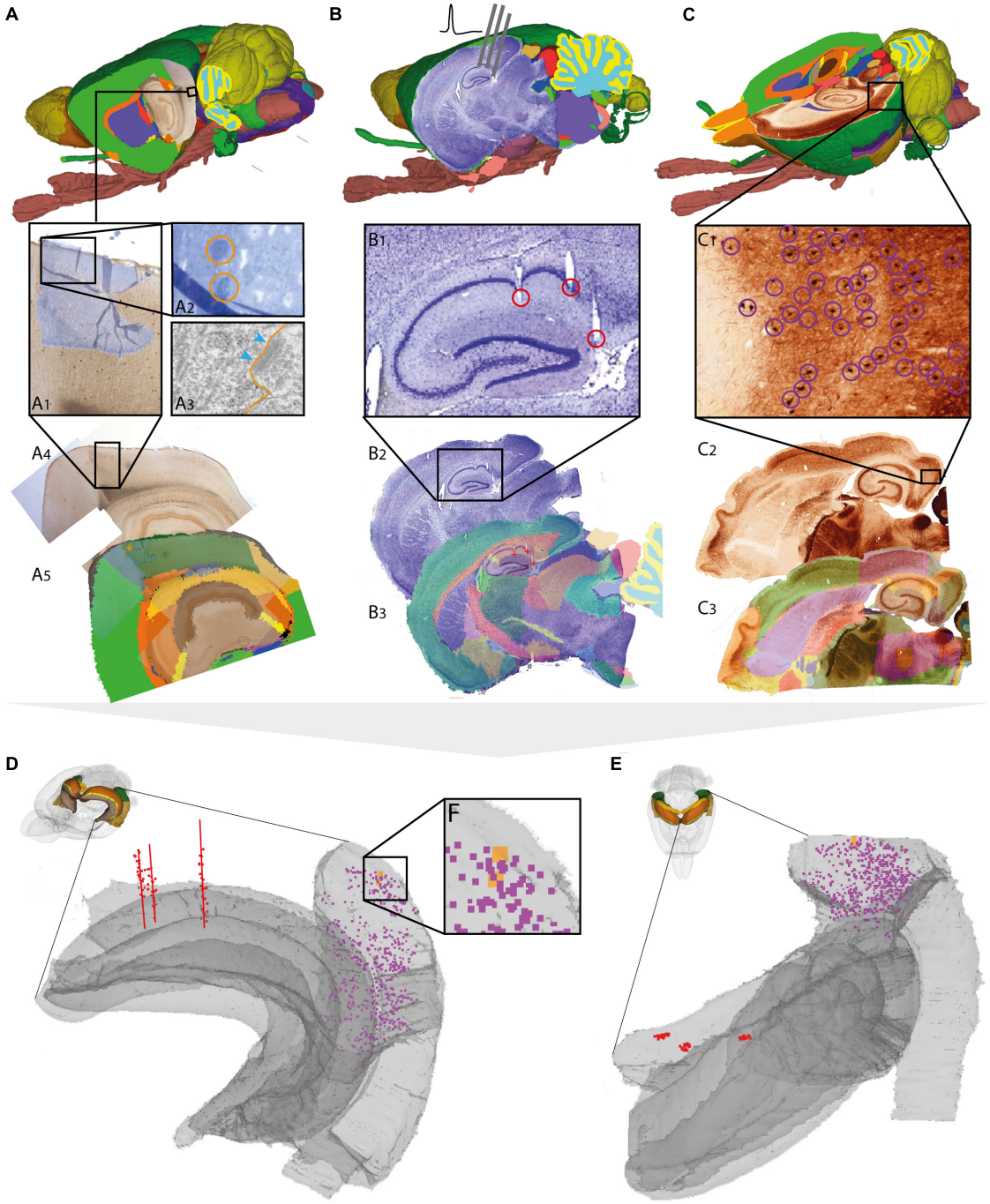


FIGURE 3 | Assigning anatomical location and integrating data using 3-D atlas and new workflows: Examples showing rat brain data. **(A–C)** Data from experimental studies using **(A)** transmission electron microscopy (Berggaard et al., 2018), **(B)** *in vivo* electrophysiology (Puchades et al., 2017), and **(C)** histochemical visualization of chemoarchitecture (Boccara et al., 2015), are mapped to the Waxholm Space atlas of the rat brain (v2, Papp et al., 2014; Kjonigsen et al., 2015). **(A1–A5)** Show the stepwise procedure used to define the location of an electron microscopy (EM) image in atlas space, by mapping an image of semithin, toluidine blue-stained section onto a low-power image showing a larger part of the brain **(A1)**. In this way the location of the parvalbumin stained cell (encircled in **A2**) shown in the EM image **(A3)** is determined in different images. By mapping the overview image **(A4)** in atlas space **(A5)**, the location of the paravalbumin positive cells shown in **(A3)** can be defined by 3-D atlas coordinates (orange crosses in **A5**). **(B1–B3)** Show how tracks of recording electrodes (encircled in **B1**), visible in a thionine stained section cut obliquely halfway between the coronal and sagittal plane **(B2)**, are mapped in atlas space by registration to the reference atlas (red crosses in **B3**). **(C1–C3)** show how positions of parvalbumin positive cells (encircled in **C1**), visualized by immunolabelling of horizontal rat brain sections **(C2)** can be determined by spatially registering section images to a reference atlas **(C3)**. **(D,E)** 3-D co-visualization of point coordinates extracted from the three data sets (**A**, orange dots; **B**, red dots, **C**, purple dots) together with gray surfaces of the right hippocampal and entorhinal regions, shown from an anterolateral **(D)** and dorsal **(E)** view. **(F)** Magnified view of data points representing the location of the cells shown in example **(A)**.

made reference atlas maps and allows extraction of spatial coordinates representing features of interest. The 3-D viewer tool MeshView was used to co-visualize the color coded coordinates from different data sets together with selected elements from the 3-D reference atlas. With this core workflow as a basis, we identified specific strategies for determining anatomical location and extracting spatial coordinates for features of interest from each methodological category. The step-wise implementation of these workflows are exemplified below for different types of data, illustrating how descriptions of location can be improved and used for data integration purposes with relatively simple steps.

For *electron microscopy* data, spatial coordinates were obtained for two parvalbumin positive cells from the medial entorhinal cortex, imaged under a transmission electron microscope (**Figure 3A3**). The ultrathin sections used for electron microscopy were sectioned from a small tissue sample dissected from a sagittal vibratome rat brain section from the temporal cortex, stained for parvalbumin by immunohistochemistry. Prior to ultrathin sectioning, semithin sections were obtained and counterstained using toluidine blue (Berggaard et al., 2018) for orientation and identification of immunopositive cells. To determine the location of the cells viewed by electron microscopy in a 3-D reference atlas, three main steps were followed (**Figures 3A1–3A5**). First, an image of the entire sagittal brain section taken prior to removal of the tissue sample was mapped to the reference atlas using QuickNII (**Figure 3A5**). Secondly, a transparent image of the semithin, toluidine blue-stained section was manually registered to the larger image of the sagittal section, by aligning specific features visible in both images, including blood vessels, labeled cells, outer surface and boundary between gray and white matter (**Figure 3A1**). Finally, the location of the parvalbumin positive cells was identified both in the semithin and ultrathin sections, and coordinates were extracted from the vibratome section (**Figure 3A5**), thus allowing identification of cells across all spatial scales. The above procedure can in principle be applied to any method involving small tissue samples, such as immunoblotting and related methods.

For *electrophysiological recording* data, spatial coordinates were extracted from the bottom of individual electrode track throughout a series of sections cut in a non-standard plane and stained to reveal cytoarchitecture (**Figure 3B**; Puchades et al., 2017). While the location of electrophysiological recordings is usually reported by use of perioperatively determined stereotaxic coordinates (**Table 1**; see above), a key step to improve precision is to determine the location of electrode tracks in histological sections. In our example, a non-standard oblique section plane was used to identify electrode tracks, which is very difficult to compare with a traditional 2-D atlas framework. Using the QuickNII tool, the section image could nevertheless be mapped to atlas space, thus allowing the location of electrode tracks to be annotated and visualized (**Figures 3B1–3B3**; Puchades et al., 2017; see also similar example shown in Bjerke et al., 2018).

For *histological material* used in microscopic studies of brain architecture, the strategy for extracting coordinates for labeled features of interest is straightforward compared to the examples above. In our example (**Figure 3C**), we used images of serial histological sections immunostained for parvalbumin (Boccara et al., 2015). After mapping the serial section images to the reference atlas, we recorded point coordinates representing immunopositive cells located within the medial entorhinal cortex in sections sampled at 200 μm intervals through the entire left entorhinal cortex (**Figure 3C1**). A similar example is shown in **Figure 4A**, using section images (downloaded from the Allen Institute for Brain Science, Lein et al., 2007)[5] showing parvalbumin positive cells visualized by *in situ* hybridization. The spatial registration of these images to the Allen Mouse Common Coordinate Framework was adjusted using the QuickNII tool, and point coordinates representing parvalbumin positive cells in the left caudoputamen were extracted from all sections (**Figure 4A2**).

For *neuron reconstructions*, a slightly different approach was used. The data included coordinate lists created by 3-D reconstruction of neurons (intracellularly filled with neurobiotin) using the Neurolucida software tool (MBF Bioscience, Williston, VT, United States), together with low-power images of sagittal sections images in which the labeled somata were visible. The sagittal section images were registered to the mouse brain atlas using QuickNII, following which the atlas coordinates corresponding to the center of the neuronal soma (seen in the histological section, cf. **Figure 4B2**) were extracted using LocaliZoom. Having determined the center point of the soma and the position and orientation of the histological section image in atlas space, we spatially translated the local (Neurolucida) coordinates representing the complete neuronal arbors of the 3-D reconstructed cell to atlas coordinates.

Thus, by mapping very different types of data to a common anatomical reference atlas, it became possible to extract point coordinates for key data features and co-visualize these in atlas space (**Figures 3D,E**, **4C–E**).

Improving Location Metadata Using New Tools and Workflows for Anatomical Localization

Based on the strategies and documentation elements used above to connect different types of data to a common anatomical framework, and extending on the minimum practice recommendations proposed above, we suggest the following additional method-independent documentation steps to improve precision and facilitate data integration: (1) document features of interest in relation to cellular or regional characteristics; (2) register images to a 3-D reference atlas framework; and, if possible (3) acquire multiple serial histological (or tomographical) images covering several anatomical landmarks.

Document Features of Interest in Relation to Cellular or Regional Characteristics

The anatomical boundaries of brain regions are usually defined by characteristic structural or functional features. These can be visualized by (immuno-)histological staining, such as the thionine or parvalbumin staining shown in **Figure 3**, or other cellular properties such as autoradiographic visualization of

[5]http://mouse.brain-map.org/experiment/show/75457579

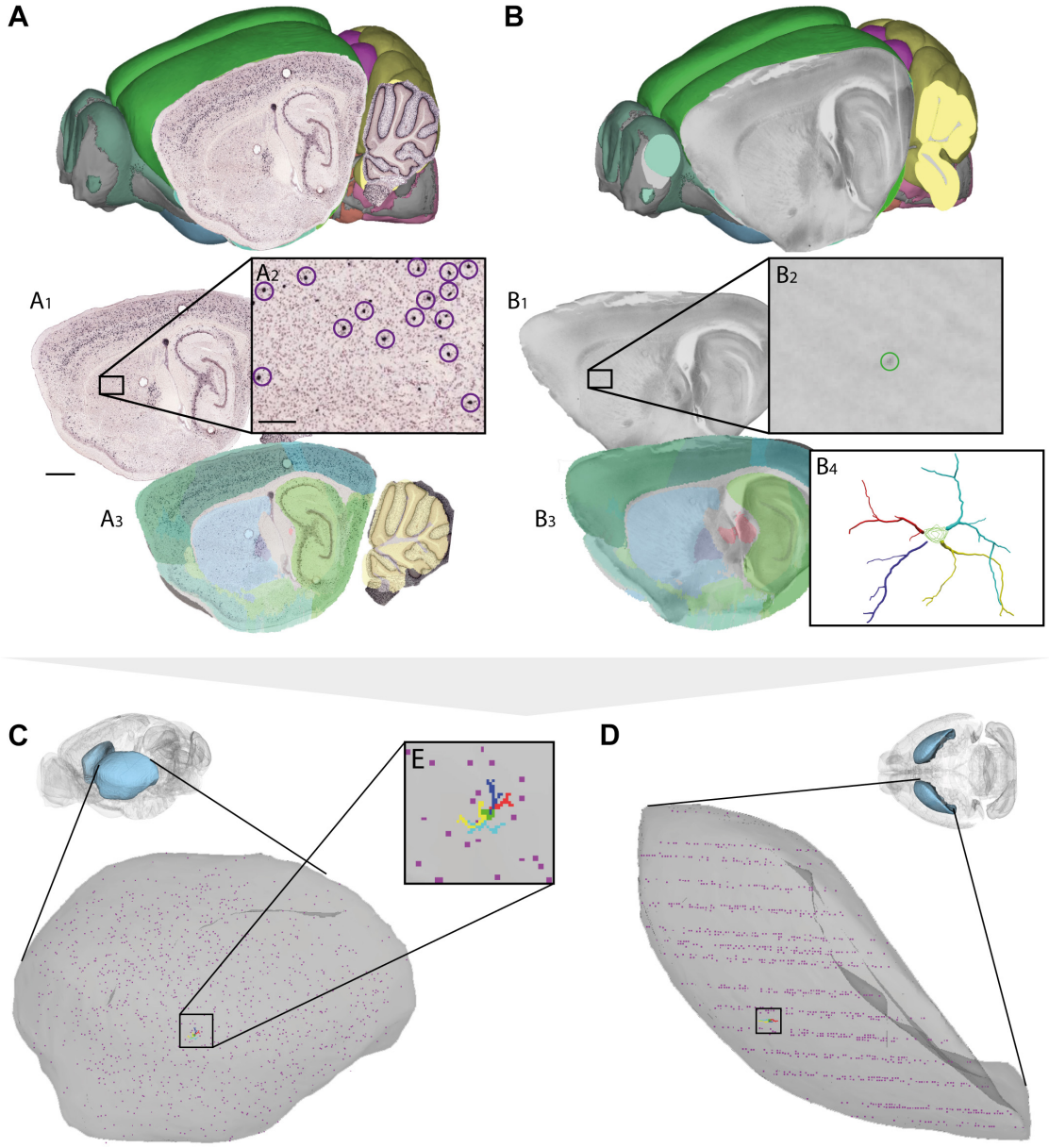


FIGURE 4 | Assigning anatomical location and integrating data using 3-D atlas and new workflows: Examples showing mouse brain data. **(A,B)** Data from experiments using **(A)** *in situ* hybridization (Lein et al., 2007) and **(B)** *in vitro* electrophysiology with single-cell reconstruction (Kozlov et al., unpublished data) mapped to the Allen Mouse Common Coordinate Framework. **(A1–A3,B1–B4)** show the stepwise procedures used to map overview images covering sagittal sections **(A1,B1)**, showing labeled cells of interest (encircled in **A2,B2**), to reference atlas space **(A3,B3)**. **(B4)** shows the 3-D reconstruction of the neuron shown in **(B2)**. **(C,D)** 3-D co-visualization of point coordinates extracted from the two data sets (**A**, purple dots; **B**, red, green, blue, and yellow dots) together with the surface of the caudoputamen, shown in view from anterolateral **(C)** and dorsal **(D)**. **(E)** Magnified view of data points representing the location of the 3-D cell reconstruction shown in example **(B)**. Scale bars: 1 mm **(A1)**, 200 μm **(A2)**.

receptors (Schubert et al., 2016) or enzyme based visualization of chemical properties (e.g., patches of cytochrome oxidase positive cell groups in the sensory whisker barrel cortex; Land and Simons, 1985). Additional approaches used to pinpoint anatomical location include, e.g., visualization of specific well-known cellular architectures or connectivity, electrophysiological measures of sensory receptive fields (Chapin and Lin, 1984) or motor-related activity (Neafsey et al., 1986), or any combination of the above. Use of such measurements can allow more fine-grained and precise anatomical descriptions of location.

Spatial Registration of Images to Reference Atlas

As shown in the examples provided in **Figures 3**, **4**, spatial registration of brain images to an anatomical atlas provides

specific evidence of location in a standard anatomical reference space. If images have been acquired with orientations matching the standard coronal, sagittal and horizontal planes used in reference atlases, such registration can simply be done by mapping 2-D diagrams from any standard reference atlases on section images. However, to correct for deviations in the angle of orientations commonly seen in histological sections, and to more directly relate positions in the experimental images to a spatial 3-D reference framework, we recommend mapping images to a 3-D atlas (Lein et al., 2007; Hawrylycz et al., 2011; Oh et al., 2014; Papp et al., 2014). Depending on the properties of the experimental material used, several software tools are available for such purposes (Majka and Wójcik, 2016; Puchades et al., 2017; **Figures 3**, **4**), allowing more accurate determination of the anatomical position and section angle in experimental material (**Figure 1B**).

Use of Serial Sections

Interpretation of anatomical location, particularly for the purpose of spatial registration, can be improved with use of serial section images that display multiple anatomical landmarks. Inclusion of more sections is particularly useful when determining deviations of section angles from the standard plane. The precision of such a registration can therefore be improved by including more section images than the ones used for analysis.

Regardless of the atlas used and the methods for relating data to it, the anatomical information as extracted from the atlas (region names, coordinates of points or sections) for the entire analyzed region(s) should be clearly communicated in publications and collections of metadata.

DISCUSSION

We have here reviewed anatomical location metadata provided in recent neuroscience publications, and found considerable differences across subfields of neuroscience. We have proposed a set of method-independent, easily adopted practices (minimum requirements) that can significantly improve reproducibility of neuroanatomical locations reported in publications. Furthermore, we have shown that re-usability and integration of data can be improved with additional steps using new software tools and workflows developed through the Human Brain Project, and that these procedures are applicable to data obtained by a range of methods.

Factors contributing to inconsistency and ambiguity in location metadata included (1) variable use of reference atlases, (2) lack of specification regarding nomenclatures, terms, and definitions used, (3) limited use of coordinate-based information, and (4) use of highly magnified image material without sufficient annotation as the only graphical display of data location. The amount of location metadata found in publications depended on the methodology with which the data had been obtained, likely pointing to different approaches and traditions having evolved as common practice within subfields of neuroscience.

The minimum requirements presented here are intended to be flexible and easily applicable to any neuroscientific method. They essentially state that descriptions of locations should be complete, and precisely define the relationship of sites of interest to anatomical landmarks, by use of semantics, coordinates or graphical representations, and preferably a combination of these. Appropriate reference to a specific nomenclature and citation of the reference atlases consulted is an obvious requirement, which is easy to implement regardless of the method used, but as our results show, often overlooked. We claim that adherence to the minimum recommendations requires little additional effort by researchers, and can substantially improve the precision of anatomical descriptions and data interpretation in neuroscience publications. Our examples specify how this can be implemented for different types of data.

However, comparison of descriptions based on text, reference atlases and image material remains dependent on substantial human interpretation. The second part of our work therefore demonstrate that data obtained by several methodologies, spanning spatial and temporal scales, may be thoroughly and accurately located in space using novel tools and workflows, and that the output of these procedures can be used to co-visualize data.

The workflows tested here for mapping data to atlas space can be implemented for any neuroscience method, provided that image material showing features of interest in relation to anatomical landmarks is available. For methods where such features are readily seen in histological section images, the procedures are quite easily applicable, as seen in our examples using immunohistochemical and *in situ* hybridization material, as well as *in vivo* electrophysiological recordings. In the case of electron microscopy data, the spatial correlation of features seen at the microscopic and ultrastructural levels is essential in order to map specific objects imaged at the electron microscopic level to a reference atlas. This was achieved here using low-power overview images acquired during tissue processing and images of semithin sections stained to show cytoarchitecture. Some steps could have been improved, e.g., by imaging the whole brain section before and after sectioning, and by keeping track of the location within the ultrathin section from which the electron microscopy images were obtained. An alternative approach would be to extract coordinates representing the perimeters of the data set, e.g., the corners of a block of tissue dissected from a vibratome section and prepared for electron microscopic imaging. Whether highly specific information about the position of individual cellular elements is desirable and attainable for a data set will depend on the research question and the methods of tissue preparation. Nevertheless, our example shows that even minor additions to common protocols (e.g., acquiring images of sections from which electron microscopic samples have been dissected) can give major improvements of precision of location metadata. For neuronal reconstruction data, we show that spatial coordinates recorded with a 3-D reconstruction software can be translated to atlas coordinates by using reference points visible in macroscopic section images. For mapping of more complex neuronal arbors in atlas space, annotation of at least four (and preferably more) reference points representing key landmarks in the neuronal reconstruction will increase precision. Again, access to low-power overview images documenting soma locatio

relative to visible landmarks was critical for translating the location of reconstructed neurons to atlas space.

We lastly summarized the workflows developed through these examples as a set of improved practices, aimed to facilitate efforts to compare and integrate neuroscience data. Mapping data to a common anatomical framework is an effective means to allow comparison and facilitate integration of disparate data types, a key goal within the Human Brain Project (Bjerke et al., 2018). Adherence to the improved practice recommendations proposed here ensures that heterogeneous data can be organized and shared in databases, with location metadata suitable for conducting queries based on location, either by using semantic strings and anatomical ontologies or by use of more fine-grained 3-D spatial queries for coordinate locations in atlas space.

Additional documentation and more extensive interpretation of anatomical locations, as exemplified above, requires additional efforts including production of additional material and documentation, as well as analytical efforts. Depending on the size of the data set, type and quality of images and the features to be extracted, the process of registering data, extracting and visualizing coordinates requires from a couple of days to a week. The workflows used to extract spatial coordinates for different data features in our examples, were based on manual annotations performed with the tool LocaliZoom. The advantage of this approach is that atlas coordinates are directly exported, but it can be tedious to apply to larger data sets. New tools and workflows are currently being developed in the Human Brain Project that will allow (semi-)automated extraction of labeled features from serial images (Kreshuk et al., 2014; Papp et al., 2016; Yates et al., 2017).

We argue that costs of such additional efforts are outweighed by improved precision of anatomical location metadata, and the added value gained by making data easier to compare across studies. Today, finding and comparing data in the literature based on a region of interest is a time-consuming task that often reveals inconsistencies in results. Indeed, flexible use of definitions has been related to poor reproducibility in science (Ioannidis, 2005). Concepts of brain regions are examples of such fluid definitions (Van De Werd and Uylings, 2014), and inaccurate reporting of location is likely to amplify the challenge caused by these changes. However, the coordinate systems that embed concepts of brain regions are static. Mapping current data to such coordinate frameworks will make data more robust in the face of evolving concepts of brain regions and is thus necessary to ensure long-term relevance of findings. Furthermore, following the improved practice recommendations outlined here can facilitate data integration and re-use of data, as the output of spatial registration procedures is structured metadata about anatomical locations that can accompany data to be shared. We therefore consider the benefits of performing these methods to outweigh the costs in the long term. New practices for data sharing in neuroscience (Ferguson et al., 2014; Leitner et al., 2016; Ascoli et al., 2017) will likely lead to augmented focus on high-quality metadata as a tool for increasing the value and impact of data, and thus also establish more prominent short-term incentives for mapping data to reference atlas space.

DATA AND SOFTWARE AVAILABILITY STATEMENT

The data sets generated for the literature survey of this study are available from the corresponding author upon request. The data sets used for testing procedures for spatial registration were obtained from public sources (Allen Institute for Brain Science repository, the Human Brain Project), or from generous colleagues as specified in the methods section. Requests to access these data sets can be directed to t.b.leergaard@medisin.uio.no. Software tools are available from the Human Brain Project (www.humanbrainproject.eu), requests for access to these tools can be directed to j.g.bjaalie@medisin.uio.no.

AUTHOR CONTRIBUTIONS

IB contributed to conceiving the study, performed analyses, contributed to development of workflows, and co-authored the manuscript. MØ and KA contributed to conceiving the study, performed analyses, and contributed to development of workflows. CB, HK, SY, and MP contributed to the analyses and development of workflows. JB supervised development of tools, infrastructure, and workflows and contributed to writing the paper. TL conceived and supervised the study, and co-authored the manuscript. All the authors reviewed and approved the manuscript.

FUNDING

This research has received funding from the European Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreement No. 720270 (Human Brain Project SGA1) and Specific Grant Agreement No. 785907 (Human Brain Project SGA2). Funding was also received from The Research Council of Norway under Grant Agreement No. 269774 (INCF National Node).

ACKNOWLEDGMENTS

We thank Gergely Csucs, Dmitri Darine, Hong Qu, and Grazyna Babinska for expert technical assistance, and Hilde Flaatten and Leiv Sandvik for helpful bibliographical advice. Histological section images were acquired at the Norbrain Slidescanning Facility at the Institute of Basic Medical Sciences, University of Oslo.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnana.2018.00082/full#supplementary-material

REFERENCES

Akhter, F., Haque, T., Sato, F., Kato, T., Ohara, H., Fujio, T., et al. (2014). Projections from the dorsal peduncular cortex to the trigeminal subnucleus caudalis (medullary dorsal horn) and other lower brainstem areas in rats. *Neuroscience* 266, 23–37. doi: 10.1016/j.neuroscience.2014.01.046

Amari, S.-I., Beltrame, F., Bjaalie, J. G., Dalkara, T., De Schutter, E., Egan, G. F., et al. (2002). Neuroinformatics: the integration of shared databases and tools towards integrative neuroscience. *J. Integr. Neurosci.* 1, 117–128. doi: 10.1142/S0219635202000128

Amunts, K., Ebell, C., Muller, J., Telefont, M., Knoll, A., and Lippert, T. (2016). The human brain project: creating a European research infrastructure to decode the human brain. *Neuron* 92, 574–581. doi: 10.1016/J.NEURON.2016.10.046

Amunts, K., Hawrylycz, M. J. J., Van Essen, D. C. C., Van Horn, J. D. D., Harel, N., Poline, J.-B. B., et al. (2014). Interoperable atlases of the human brain. *Neuroimage* 99, 525–532. doi: 10.1016/j.neuroimage.2014.06.010

Ascoli, G. A., Maraver, P., Nanda, S., Polavaram, S., and Armananaz, R. (2017). Win-win data sharing in neuroscience. *Nat. Methods* 14, 112–116. doi: 10.1038/nmeth.4152

Berggaard, N., Bjerke, I. E., Paulsen, A. E. B., Hoang, L., Skogaker, N. E. T., Menno, P., et al. (2018). Development of Parvalbumin-expressing basket terminals in layer II of the rat medial entorhinal cortex. *eNeuro* 5:ENEURO.0438-17.2018 doi: 10.1523/ENEURO.0438-17.2018

Bjaalie, J. G. (2002). Opinion: localization in the brain: new solutions emerging. *Nat. Rev. Neurosci.* 3, 322–325. doi: 10.1038/nrn790

Bjaalie, J. G. (2008). Understanding the brain through neuroinformatics. *Front. Neurosci.* 2, 19–21. doi: 10.3389/neuro.01.022.2008

Bjaalie, J. G., Leergaard, T. B., Lillehaug, S., Odeh, F., Moene, I. A., Kjode, J. O., et al. (2005). Database and tools for analysis of topographic organization and map transformations in major projection systems of the brain. *Neuroscience* 136, 681–695. doi: 10.1016/j.neuroscience.2005.06.036

Bjerke, I. E., Øvsthus, M., Papp, E. A., Yates, S. C., Silvestri, L., Fiorilli, J., et al. (2018). Data integration through brain atlasing: human brain project tools and strategies. *Eur. Psychiatry* 50, 70–76. doi: 10.1016/j.eurpsy.2018.02.004

Boccara, C. N., Kjonigsen, L. J., Hammer, I. M., Bjaalie, J. G., Leergaard, T. B., and Witter, M. P. (2015). A three-plane architectonic atlas of the rat hippocampal region. *Hippocampus* 25, 838–857. doi: 10.1002/hipo.22407

Boline, J., Lee, E.-F., and Toga, A. W. (2008). Digital atlases as a framework for data sharing. *Front. Neurosci.* 2, 100–106. doi: 10.3389/neuro.01.012.2008

Boy, J., Leergaard, T. B., Schmidt, T., Odeh, F., Bichelmeier, U., Nuber, S., et al. (2006). Expression mapping of tetracycline-responsive prion protein promoter: digital atlasing for generating cell-specific disease models. *Neuroimage* 33, 449–462. doi: 10.1016/j.neuroimage.2006.05.055

Chapin, J. K., and Lin, C.-S. (1984). Mapping the body representation in the SI cortex of anesthetized and awake rats. *J. Comp. Neurol.* 229, 199–213. doi: 10.1002/cne.902290206

Dobi, A., Sartori, S. B., Busti, D., Van Der Putten, H., Singewald, N., Shigemoto, R., et al. (2013). Neural substrates for the distinct effects of presynaptic group III metabotropic glutamate receptors on extinction of contextual fear conditioning in mice. *Neuropharmacology* 66, 274–289. doi: 10.1016/j.neuropharm.2012.05.025

Ferguson, A. R., Nielson, J. L., Cragin, M. H., Bandrowski, A. E., and Martone, M. E. (2014). Big data from small data: data-sharing in the "long tail" of neuroscience. *Nat. Neurosci.* 17, 1442–1447. doi: 10.1038/nn.3838

Hawrylycz, M., Baldock, R. A., Burger, A., Hashikawa, T., Johnson, G. A., Martone, M., et al. (2011). Digital atlasing and standardization in the mouse brain. *PLoS Comput. Biol.* 7:e1001065. doi: 10.1371/journal.pcbi.1001065

Hey, T., and Trefethen, A. (2003). "The data deluge: an e-science perspective," in *Grid Computing - Making the Global Infrastructure a Reality*, eds F. Berman, G. Fox, and T. Hey (Hoboken, NJ: Wiley), 809–824.

Hintiryan, H., Gou, L., Zingg, B., Yamashita, S., Lyden, H. M., Song, M. Y., et al. (2012). Comprehensive connectivity of the mouse main olfactory bulb: analysis and online digital atlas. *Front. Neuroanat.* 6:30. doi: 10.3389/fnana.2012.00030

Hjornevik, T., Leergaard, T. B., Darine, D., Moldestad, O., Dale, A. M., Willoch, F., et al. (2007). Three-dimensional atlas system for mouse and rat brain imaging data. *Front. Neuroinform.* 1:4. doi: 10.3389/neuro.11.004.2007

Insausti, R., Herrero, M. T., and Witter, M. P. (1997). Entorhinal cortex of the rat: cytoarchitectonic subdivisions and the origin and distribution of cortical efferents. *Hippocampus* 7, 146–183. doi: 10.1002/(SICI)1098-1063(1997)7:2<146::AID-HIPO4>3.0.CO;2-L

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Med.* 2:e124. doi: 10.1371/journal.pmed.0020124

Johnson, G. A., Badea, A., Brandenburg, J., Cofer, G., Fubara, B., Liu, S., et al. (2010). Waxholm space: an image-based reference for coordinating mouse brain research. *Neuroimage* 53, 365–372. doi: 10.1016/J.NEUROIMAGE.2010.06.067

Kjonigsen, L. J., Leergaard, T. B., Witter, M. P., and Bjaalie, J. G. (2011). Digital atlas of anatomical subdivisions and boundaries of the rat hippocampal region. *Front. Neuroinform.* 5:2. doi: 10.3389/fninf.2011.00002

Kjonigsen, L. J., Lillehaug, S., Bjaalie, J. G., Witter, M. P., and Leergaard, T. B. (2015). Waxholm Space atlas of the rat brain hippocampal region: three-dimensional delineations based on magnetic resonance and diffusion tensor imaging. *Neuroimage* 108, 441–449. doi: 10.1016/j.neuroimage.2014.12.080

Koslow, S. H., and Subramaniam, S. (eds). (2005). *Databasing the Brain: From Data to Knowledge*. New York, NY: John Wiley & Sons, Inc.

Kreshuk, A., Koethe, U., Pax, E., Bock, D. D., and Hamprecht, F. A. (2014). Automated detection of synapses in serial section transmission electron microscopy image stacks. *PLoS One* 9:e87351. doi: 10.1371/journal.pone.0087351

Land, P. W., and Simons, D. J. (1985). Cytochrome oxidase staining in the rat smI barrel cortex. *J. Comp. Neurol.* 238, 225–235. doi: 10.1002/cne.902380209

Lein, E. S., Hawrylycz, M. J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., et al. (2007). Genome-wide atlas of gene expression in the adult mouse brain. *Nature* 445, 168–176. doi: 10.1038/nature05453

Leitner, F., Bielza, C., Hill, S. L., and Larrañaga, P. (2016). Data publications correlate with citation impact. *Front. Neurosci.* 10:419. doi: 10.3389/fnins.2016.00419

Majka, P., and Wójcik, D. K. (2016). Possum—a framework for three-dimensional reconstruction of brain images from serial sections. *Neuroinformatics* 14, 265–278. doi: 10.1007/s12021-015-9286-1

Neafsey, E. J., Bold, E. L., Haas, G., Hurley-Gius, K. M., Quirk, G., Sievert, C. F., et al. (1986). The organization of the rat motor cortex: a microstimulation mapping study. *Brain Res.* 396, 77–96. doi: 10.1016/0165-0173(86)90011-1

Oh, S. W., Harris, J. A., Ng, L., Winslow, B., Cain, N., Mihalas, S., et al. (2014). A mesoscale connectome of the mouse brain. *Nature* 508, 207–214. doi: 10.1038/nature13186

Papp, E. A., Leergaard, T. B., Calabrese, E., Johnson, G. A., and Bjaalie, J. G. (2014). Waxholm space atlas of the sprague dawley rat brain. *Neuroimage* 97, 374–386. doi: 10.1016/j.neuroimage.2014.04.001

Papp, E. A., Leergaard, T. B., Csucs, G., and Bjaalie, J. G. (2016). Brain-wide mapping of axonal connections: workflow for automated detection and spatial analysis of labeling in microscopic sections. *Front. Neuroinform.* 10:11. doi: 10.3389/fninf.2016.00011

Paxinos, G., and Franklin, K. (2012). *The Mouse Brain in Stereotaxic Coordinates*, 4th Edn, San Diego, CA: Academic Press.

Paxinos, G., and Watson, C. (2005). *The Rat Brain in Stereotaxic Coordinates*, 5th Edn, San Diego, CA: Elsevier.

Paxinos, G., and Watson, C. (2013). *The Rat Brain in Stereotaxic Coordinates*, 7th Edn, Burlington, NJ: Elsevier Inc.

Puchades, M. A., Csucs, G., Checinska, M., Øvsthus, M., Bjerke, I. E., Andersson, K., et al. (2017). *QuickNII: Neuroinformatics Tool and Workflow for Anchoring of Serial Histological Images in Rodent Brain 3D Space. Abstract no. 532.12 in Neuroscience Meeting Planner*. Washington, DC: Society for Neuroscience.

Ragan, T., Kadiri, L. R., Venkataraju, K. U., Bahlmann, K., Sutin, J., Taranda, J., et al. (2012). Serial two-photon tomography for automated ex vivo mouse brain imaging. *Nat. Methods* 9, 255–258. doi: 10.1038/nmeth.1854

Schubert, N., Axer, M., Schober, M., Huynh, A.-M., Huysegoms, M., Palomero-Gallagher, N., et al. (2016). 3D reconstructed Cyto-, Muscarinic M2 receptor, and fiber architecture of the rat brain registered to the waxholm space atlas. *Front. Neuroanat.* 10:51. doi: 10.3389/fnana.2016.00051

Sergejeva, M., Papp, E. A., Bakker, R., Gaudnek, M. A., Okamura-Oho, Y., Boline, J., et al. (2015). Anatomical landmarks for registration of experimental image data to volumetric rodent brain atlasing templates. *J. Neurosci. Methods* 240, 161–169. doi: 10.1016/j.jneumeth.2014.11.005

Stopps, M., Allen, N., Barrett, R., Choudhury, H. I., Jarolimek, W., Johnson, M., et al. (2004). Design and application of a novel brain slice system that permits independent electrophysiological recordings from multiple slices. *J. Neurosci. Methods* 132, 137–148. doi: 10.1016/j.jneumeth.2003.08.015

Swanson, L. (2004). *Brain Maps: Structure of the Rat Brain*, 3rd Edn, Amsterdam: Elsevier.

Tiesinga, P., Bakker, R., Hill, S., and Bjaalie, J. G. (2015). Feeding the human brain model. *Curr. Opin. Neurobiol.* 32, 107–114. doi: 10.1016/j.conb.2015.02.003

Van De Werd, H. J., and Uylings, H. B. (2014). Comparison of (stereotactic) parcellations in mouse prefrontal cortex. *Brain Struct. Funct.* 219, 433–459. doi: 10.1007/s00429-013-0630-7

Veraart, J., Leergaard, T. B., Antonsen, B. T., Van Hecke, W., Blockx, I., Jeurissen, B., et al. (2011). Population-averaged diffusion tensor imaging atlas of the Sprague Dawley rat brain. *Neuroimage* 58, 975–983. doi: 10.1016/j.neuroimage.2011.06.063

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3, 1–9. doi: 10.1038/sdata.2016.18

Witter, M. (2012). "Hippocampus," in *The Mouse Nervous System*, ed. L. Puelles (Amsterdam: Elsevier), 112–139. doi: 10.1016/B978-0-12-369497-3.10005-6

Yates, S. C., Puchades, M. A., Coello, C., Kreshuk, A., Hartlage-Rübsamen, M., Rossner, S., et al. (2017). *Workflow for Automated Quantification and Spatial Analysis of Labeling in Microscopic Rodent Brain Sections. Abstract no. 44.15 in Neuroscience Meeting Planner*. Washington, DC: Society for Neuroscience.

Zakiewicz, I. M., van Dongen, Y. C., Leergaard, T. B., and Bjaalie, J. G. (2011). Workflow and atlas system for Brain-Wide mapping of axonal connectivity in rat. *PLoS One* 6:e22669. doi: 10.1371/journal.pone.0022669

Zaslavsky, I., Baldock, R. A., and Boline, J. (2014). Cyberinfrastructure for the digital brain: spatial standards for integrating rodent brain atlases. *Front. Neuroinform.* 8:74. doi: 10.3389/fninf.2014.00074

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

II

SCIENTIFIC DATA

OPEN

DATA DESCRIPTOR

Database of literature derived cellular measurements from the murine basal ganglia

Ingvild E. Bjerke, Maja A. Puchades, Jan G. Bjaalie & Trygve B. Leergaard ✉

Quantitative measurements and descriptive statistics of different cellular elements in the brain are typically published in journal articles as text, tables, and example figures, and represent an important basis for the creation of biologically constrained computational models, design of intervention studies, and comparison of subject groups. Such data can be challenging to extract from publications and difficult to normalise and compare across studies, and few studies have so far attempted to integrate quantitative information available in journal articles. We here present a database of quantitative information about cellular parameters in the frequently studied murine basal ganglia. The database holds a curated and normalised selection of currently available data collected from the literature and public repositories, providing the most comprehensive collection of quantitative neuroanatomical data from the basal ganglia to date. The database is shared as a downloadable resource from the EBRAINS Knowledge Graph (https://kg.ebrains.eu), together with a workflow that allows interested researchers to update and expand the database with data from future reports.

Background & Summary

Quantitative knowledge about the number and normal variation of different cell types of the brain and their subcellular elements, such as synapses and dendritic spines, is of broad interest for neuroscientists. This is important for several purposes, including building and constraining computational models[1–3], guiding new experimental research, and comparing data from individual or groups of subjects. The need for quantitative measurements of neural architecture has led to development of numerous experimental methods for unbiased quantification of neuroanatomical features. Examples include cell counting methods[4], stereological approaches to obtain numbers, areas or volumes[5,6], and point pattern analyses to characterise spatial distributions of cells or cellular elements[7,8]. The results of such studies are typically published in original papers, reporting e.g. estimates of total numbers or densities of cells[9,10] or relative amounts of cells, synapses, spines, or other parameters in different experimental groups[11,12]. While individual papers may be easily interpreted, it is becoming increasingly challenging to overview the steadily growing amount of publications[13] and to evaluate the consistency and comparability of information. The traditional research paper format is not particularly well suited to make comparisons, as data may be distributed across text, tables and figures, with units of measurements and nomenclatures that vary across papers. Although units of measurements can be effectively converted and nomenclature differences may be possible to resolve, this requires significant time and effort from the reader. In some cases, findings are reported in non-standard units (e.g. as percentage of control, number per section), which may make them impossible to compare to other results. Researchers investigating brain structure and function in animal models may find it difficult to answer relatively simple questions, such as: What is the average number of cells or subcellular structures in my brain region of interest, and how much do these numbers vary? Which parameters have been quantified before and what were the methods used to do so? Can data from two studies be compared? Are the results reported in the literature within the same range?

Neuroanatomical information is available from several databases. The temporal-lobe database (www.temporal-lobe.com) presents connections in the rat hippocampal region are presented schematically in an interactive PDF, allowing the user to quickly view aggregated information[14]. In the Brain Architecture Management System (BAMS, https://bams1.org/) project, Bota and colleagues compiled reports and scored the strength of connections between regions across the brain on a semi-quantitative scale[15]. The hippocampome (www.

Department of Molecular Medicine, Institute of Basic Medical Sciences, University of Oslo, Oslo, Norway. ✉e-mail: t.b.leergaard@medisin.uio.no

hippocampome.org), a database of neuronal cell types in the hippocampus, contains interactive matrices showing the location, cytochemistry, electrophysiology, and connectivity of different cell types[16]. The NeuroMorpho database[17] (www.neuromorpho.org) is an extensive collection of published neuron morphologies, with useful quantitative information about the features of individual neurons. In addition, several efforts have been made to estimate brain cell numbers in histological material using computational image analysis[18–20]. To our knowledge, no systematic effort has been made to collect and normalise information from several sources about the number and distribution of cells, synapses and spines in different brain regions in a database.

We here present a database of publicly available quantitative measurements of cells, synapses and dendritic spines of the frequently investigated murine basal ganglia. These are regions of high interest for basic experimental studies of voluntary movements, procedural learning and neurodegenerative diseases such as Parkinson's and Huntington's disease[21,22]. Quantitative information about the cellular architecture of these regions in normal animals is needed for computational modelling efforts[1,2], and represent an important benchmark for interpretation of results from experimental studies in different animal disease models[23,24]. The database holds > 1200 quantitative estimates derived from the literature and public repositories, normalised to standard units of measurements and mapped to common anatomical reference atlases. To our knowledge, this is the most extensive collection of available information on cellular basal ganglia parameters to date. The database is publicly shared via EBRAINS[25], together with a workflow for updating it with results from future analyses. We believe this can be a valuable benchmark resource for anatomical studies or efforts to model the murine basal ganglia.

Methods

Overview of study design. We created a database of data derived from the literature and public repositories. We here use the term derived data to describe the specific analytic results of a study, e.g. the number of cells in a given region, as opposed to the raw data that were used to generate this number. We limited our scope to quantitative information about number, distribution and morphology of cells and subcellular elements of the normal, rodent basal ganglia. We here consider the concept of the basal ganglia to include dorsal and ventral parts of the striatum (caudoputamen and nucleus accumbens) and pallidum (external globus pallidus, entopeduncular nucleus and ventral pallidum), as well as the subthalamic nucleus and substantia nigra[26,27]. We designed a database using Microsoft Access, and set up a search string to query the literature. Specific inclusion criteria were used to narrow the number of papers to include. For each paper, the methods and results sections were carefully read and annotated, and relevant metadata elements were integrated in the database. We also searched for repositories with relevant data. Wherever necessary and possible, we standardised terms and units used to describe data, and novel workflows were developed to map data to common schemes for regions and cell types of interest. Lastly, in order to make our database accessible and usable to the community, we shared it as a dataset through the EBRAINS Knowledge Graph (RRID:SCR_017612). Each main part of the study design (database design, search strategy, data / metadata standardisation, and data sharing) will be elaborated in the following.

Database design. We organised data derived from the literature in a Microsoft Access database with 45 tables, the most important of which are summarised in Fig. 1. All fields in all tables of the database are listed and explained in Supplementary File 1.

Search strategy. *Literature search strategy*. PubMed was queried via Ovid Medline for papers from 1946-present. This search string included key words related to (1) species of interest, i.e. rat and mouse; (2) brain regions of interest, i.e. basal ganglia regions; (3) methods of interest, i.e. typical methods employed in anatomical and morphological studies; and (4) parameters of interest, i.e. numbers, densities or distributions of cells, synapses, axonal boutons, or dendritic spines. The papers needed to contain one key word from each of the categories in either the title or the abstract (the parts were combined with AND operators).

Three searches were performed, and search strings used are included in Supplementary File 2. In this iteration, we included all number, density and distribution data from the basal ganglia of adult, naïve rats or mice. However, the derived data was quite heterogeneous and few numbers could be compared. In the second and third search, we opted to include more data representing similar parameters. To this end, we narrowed the scope to data from the substantia nigra (second search) and caudoputamen (third search), but broadened the inclusion criteria to include all control animals of all (postnatal) ages.

The first search was performed on January 3rd, 2018, and a total of 2246 papers were returned. All of these papers were manually screened, and included or excluded based on a set of predefined criteria. The data had to fit with the overall criteria specified in the search string (e.g. neuroscience related, murine data, and original article format) and be available in English. Furthermore, we only included papers with data related to adult, *naïve* animals, that is, animals that had not been subject to any experimental or control manipulations, behavioural tests or training, or any other experimental intervention. The only exception to this criterion was made where pooled data from two control groups were given (e.g. sham operation and naïve control) where individual measurements had been statistically compared and proven similar. Animals with genetic manipulations, e.g. fluorescent expression in certain cells, were also excluded. Non-naïve animals were excluded to reduce the number of included publications in this first iteration of the search. However, in later and more specific queries, studies of non-naïve animals were included in order to avoid missing clearly relevant data (see below). Papers had to present *quantitative* data of interest in text or tabular format, excluding papers presenting data in graphs only. Lastly, data needed to be possible to normalise to a common unit of measurement. This generally meant that data had to be presented as numbers representing either a region of interest or a standard unit (square or cubic nano, micro-, or millimetre). In contrast, we excluded data that were presented as relative measures such as percentage of control or numbers

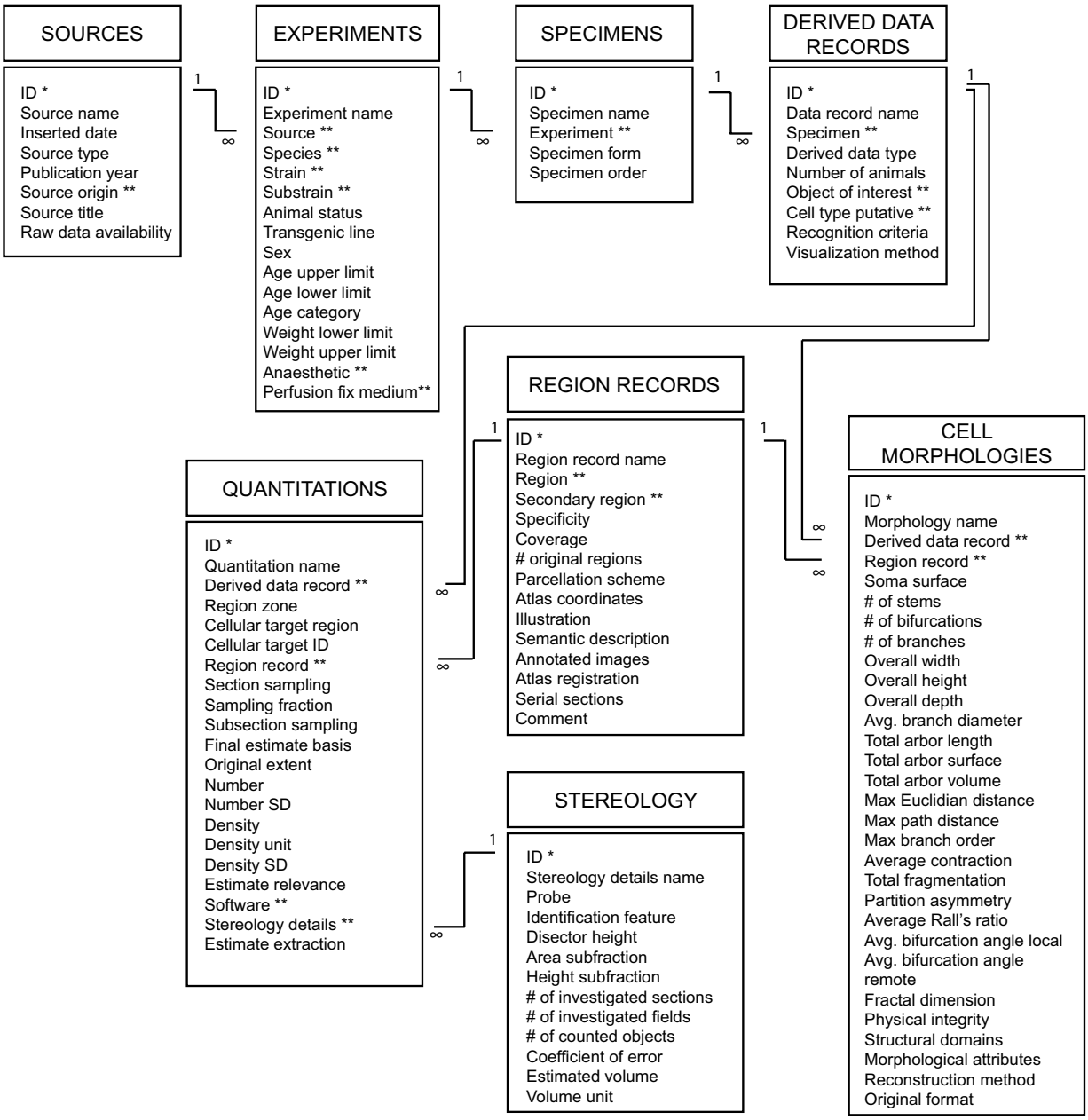


Fig. 1 Key tables in the database. Information about each source (publication or repository) is stored in the "Sources" table. Each source may have one or more related record in the "Experiments" table. An experiment may have one or more specimens, information about which is stored in the "Specimens" table. Specimens are not defined at the level of individual animals, but rather at the level of different *types* of specimen, e.g. a brain, serial sections, etc. The specimens are organised in levels, i.e. the whole brain is the primary specimen, and if a series of sections are cut from that brain it would be a secondary specimen. A specimen may have one or more derived data records, information about which is stored in the derived data records table. A derived data record can relate to one or more records in either the "Quantitations" or "Cell morphologies" table. A record in the quantitations table may or may not have a related record in the "Stereology details" table, depending on whether such procedures have been used; if similar parameters were used for stereological counting, several quantitations may relate to the same record in this table. The "Region records" table stores information about the location of one or more quantitation or morphology in standard EBRAINS atlas terms; in addition, information is stored about the accuracy of the translation from the original term used by the authors, as well as the documentation provided to support location information. For more information about each field in all tables of the database, including those not shown in this figure, see Supplementary File 1. *Primary key; **Foreign key.

per section. After manual screening, we included 65 publications with data of interest from the normal adult rat or mouse basal ganglia. An additional eight papers were included through tracking references of particularly relevant papers approach so that 72 papers were ultimately included in the first search.

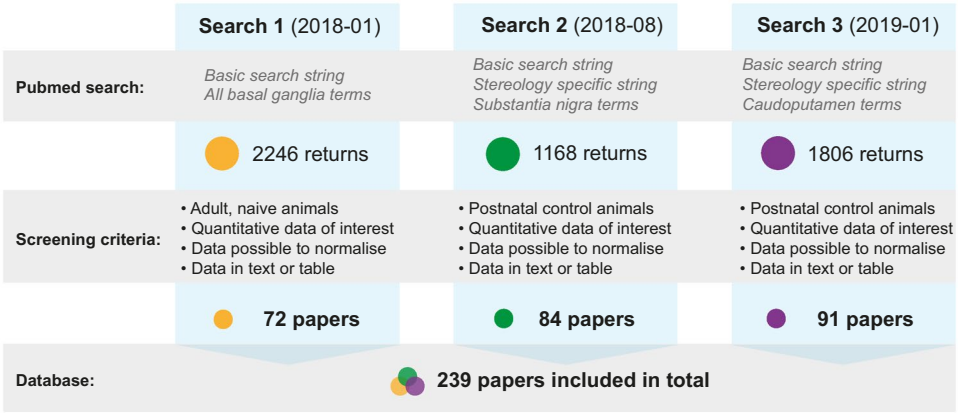


Fig. 2 Search strategy, inclusion criteria and results for the literature search. We performed three iterations of our PubMed search (search 1–3) and manually screened all returned papers to collect data of interest for the current project. The basic search string was the same in each of these iterations, but with changes in which basal ganglia regions were included. In the second and third search, we also used an additional, targeted search string to retrieve sterological studies. The full search strings are included in Supplementary file 2. In the first iteration of the search, we only included data from naïve (untreated) adult animals from all basal ganglia regions. In the second and third search, we narrowed our scope to the substantia nigra and caudoputamen, respectively, but included data from all (postnatal) control animals. See text for further details about the criteria. In the end, 239 unique papers were included in our database (some papers appeared in two or more of the searches).

The search string employed was a compromise between sensitivity and specificity, with the use of keywords related to tissue preparation method (e.g. immunohistochemistry, immunofluorescence, histology) reducing the number of search entries considerably. Since a substantial proportion of papers found with the initial search were excluded during manual screening, these keywords were included to narrow the number of papers returned. Nevertheless, to avoid missing clearly relevant papers, we performed an additional, targeted search for papers particularly conducting stereological counting in the next iterations of the search. Thus, two separate search strings were used for the second search: 1) the same string as in the first search, but with substantia nigra keywords only (performed on August 14th, 2018); and 2) an additional search string including only (stereolog*) and the keywords related to substantia nigra (performed on August 22nd, 2018). In the third search, we repeated both parts of the second search, but with the striatum (caudoputamen) as the region of interest. The two parts of the last search were performed 1) on November 30th, 2018 and 2) on January 17th, 2019. All the papers were screened manually, according to essentially the same criteria as for the first search, except that we included all control animals of all (postnatal) ages, including those genetically altered to express fluorescence in certain cells. We also included studies using animals that had been treated according to standard neuroanatomical protocols, e.g. axonal tract tracing experiments. The second search returned 1168 papers of which 84 were included, while the third search yielded 1806 papers of which 91 were included. Because some of the papers appeared in more than one of the search rounds, the total number of publications ultimately included in the database was 239. The search strategy, inclusion criteria and results for each iteration of the search is summarised in Fig. 2.

To limit the selection of papers and thus the scope of the survey, we excluded papers presenting data in graphs only. However, for studies presenting some material in text and some in graphs, we digitised graphs to extract all relevant data from the paper. We used a web-based plot digitiser (https://apps.automeris.io/wpd/) to import graph images, added reference points, and extracted the relevant means and error measurements. As this approach was quite time consuming, we used it for selected papers in the first and second search only. We included a field in the database to specify whether an estimate was extracted from text or from a graph.

Repository search strategy. Several data and metadata repositories exist with various types of neuroscience information, and the Neuroscience Information Framework (NIF, www.neuinfo.org; RRID:SCR_002894)[28] catalogues these resources. We therefore searched the NIF for portals or databases related to rat or mouse, which returned 281 public repositories with information from rats or mice. From these, we selected nine resources that appeared to be relevant to the current project. To be included, a repository had to include relevant derived data in addition to appropriate metadata. Two repositories fulfilled these criteria: Neuromorpho (www.neuromorpho.org; RRID:SCR_002145; neuronal morphology information[29], e.g. soma size, number of bifurcations; see frequently asked questions at www.neuromorpho.org for a full list) and Mouse Brain Architecture (www.brainarchitecture.org; RRID:SCR_004683; cell densities). Since the NeuroMorpho database is organised in several archives, each containing data from one laboratory, each such archive was treated as a separate source in our database, with the prefix "NMO" in the source name indicating that the data came from NeuroMorpho. A table of the evaluated repositories is included in Supplementary File 3. Lastly, we included information extracted from an Allen mouse brain *in situ* hybridisation experiment, available as a derived dataset via the EBRAINS Knowledge Graph[30].

Data and metadata standardisation. To give a unified view of data, we mapped them to key features in common schemes. Two particularly important such features in neuroscience are anatomical region and cell type

of interest. Indeed, other databases have generally been structured around regions[14] or cell types[16,31], or both[15]. In the following, we describe the workflows established in this project to map data to common terms for regions and cell types of interest, as well as how all data were standardised to common units of measurement.

Mapping data to semantically defined anatomical regions of interest. Reference atlases are commonly used in neuroscience in order to relate data to anatomical locations in the brain; however, there are several alternatives to reference atlas available just for the rat[32–35] or mouse brain[36,37], that vary with respect to how they name and define regions. Data related to a specific region in one atlas are therefore not necessarily easily compared to data related to a similarly named region in another. Even data related to the same region in different versions of the same atlas may not be directly comparable, since some borders may have been significantly revised between atlas versions.

In our database, all data were related to the three-dimensional (3D) standard atlas templates used in EBRAINS – the Waxholm space atlas of the rat brain (WHS, version 1.01[34,38]; RRID:SCR_017124) and the Allen Mouse Brain Common Coordinate Framework (CCF, version 3[37]). Using the QuickNII software for registration of 2D section images to 3D atlases[39] (RRID:SCR_016854), we mapped plates from several of the most common atlases[32,33,40–49] to WHS or CCF (Fig. 3a). The location metadata, specifying the parameters used to spatially register the different atlases to the WHS or CCF, are available as datasets from the EBRAINS Knowledge Graph[50–62]. We used the spatially co-registered atlas diagrams to inspect and define the spatial relationships between our regions of interest (basal ganglia regions) in the WHS or CCF and regions defined in other atlases. The type of relationships were categorized as *identical*, *part of, includes*, *overlapping*, or *non-overlapping* (Fig. 3c). The latter was used only in cases where regions could be expected to be related (e.g. by sharing the same name), but were found not to be. Descriptive comments about the relationships were added. In addition, to semi-quantitatively describe the degree of comparability of two regions, we applied a region comparability score, ranging from zero (non-overlapping structures) to 10 (completely identical structures). The criteria underlying this scoring system and categorization of relationships are provided in Supplementary File 4. The accumulated information about the spatial relationships defined between atlas regions are shared through the EBRAINS Knowledge Graph as separate data sets[61,62].

The locations of data presented in papers were not always defined with use of terms from a specific reference atlas. We considered data to be related to a region in an atlas only in cases where it could be clearly inferred which region (or set of regions) in the cited atlas authors referred to. This generally involved use of a specific reference to an atlas and a region name existing in that particular atlas, with a few exceptions where authors referred to a region at a lower granularity than given in the atlas. For example, although the exact term 'substantia nigra' does not appear in most atlases (the region is usually subdivided, at least into a reticular and a compact part), it is reasonable to use this term to refer to the various substructures together. The crucial point is to define the inclusion or exclusion of subdivisions *as they are named in the particular atlas*. In cases where this was not clearly defined, it was reflected in our translation by storing the coverage and specificity as "unknown". For data not defined in terms of a reference atlas, we considered the region to be defined in a 'custom' parcellation scheme. In these cases, knowledge about relations to our atlases could only be inferred from the documentation provided by the authors. To translate such custom terms, we therefore carefully considered the documentation and assigned an atlas term based on our knowledge about the basal ganglia regions and terms typically used to describe them. In general, more well-documented regions of interest allowed for more accurate translation with higher confidence.

For each mention of a region of interest, we included metadata describing how it was documented and stored this in the "Region records" table. We furthermore calculated a score to capture the degree to which each region of interest was documented (referred to as a "documentation score". Different types of documentation were weighed differently, and a score between 1 and 10 was calculated. Information about the documentation factors and their weight in the documentation score can be found under the "Region records" table section in Supplementary File 1.

Mapping data to cell types of interest. All of the objects for which we collected quantitative information in this study belong to a cell: subcellular objects originate from a cell of interest, and reconstructed and counted cells have an identity. Cell type classification is not trivial[63,64], as there are many complementary approaches to the task (e.g. cytochemical, electrophysiological, morphological), and thus no standard ontologies of cell types exist. To map data to cell types, we captured information about the various phenotypes that a cell might have. This approach was inspired by ongoing work from the INCF special interest group on Neuroinformatics for cell types (https://www.incf.org/sig/neuroinformatics-cell-types). We included seven broad phenotype categories: *brain region* (e.g. striatum, substantia nigra), *expression* (e.g. parvalbumin, tyrosine hydroxylase), *electrophysiology* (e.g. fast spiking), *morphology* (e.g. spiny neuron, giant neuron), *connectivity* (e.g. direct pathway neuron), *local connectivity* (e.g. perisomatic neuron), and *circuit function* (e.g. inhibitory neuron). For every derived data set, information was stored about the phenotype recorded for the particular cell. One or more phenotypes might be used in a particular study to classify the neuron type(s), and based on the phenotype(s) identified, a putative cell type was assigned. Some data spanned several different cell types, for example when numbers of *all* objects of interest were counted regardless of type (e.g. counting all dendritic spines or cell bodies in a certain area). In these cases, data are relevant for *all* cell types, and have simply been linked to the type "Cell", "Neuron", or "Glia", depending on the phenotypes identified.

Standardisation to common units of measurement. Prior to data entry, we converted all density units to square or cubic milli- or micrometres. Standard errors were converted to standard deviations by dividing by the square root of the sample size. Information about calculations performed to standardise a measurement was entered in the database. For data given per square milli- or micrometres, we calculated the volumetric density by dividing numbers by section thickness[65]. These were entered in the database *in addition* to original 2D counts. Numbers

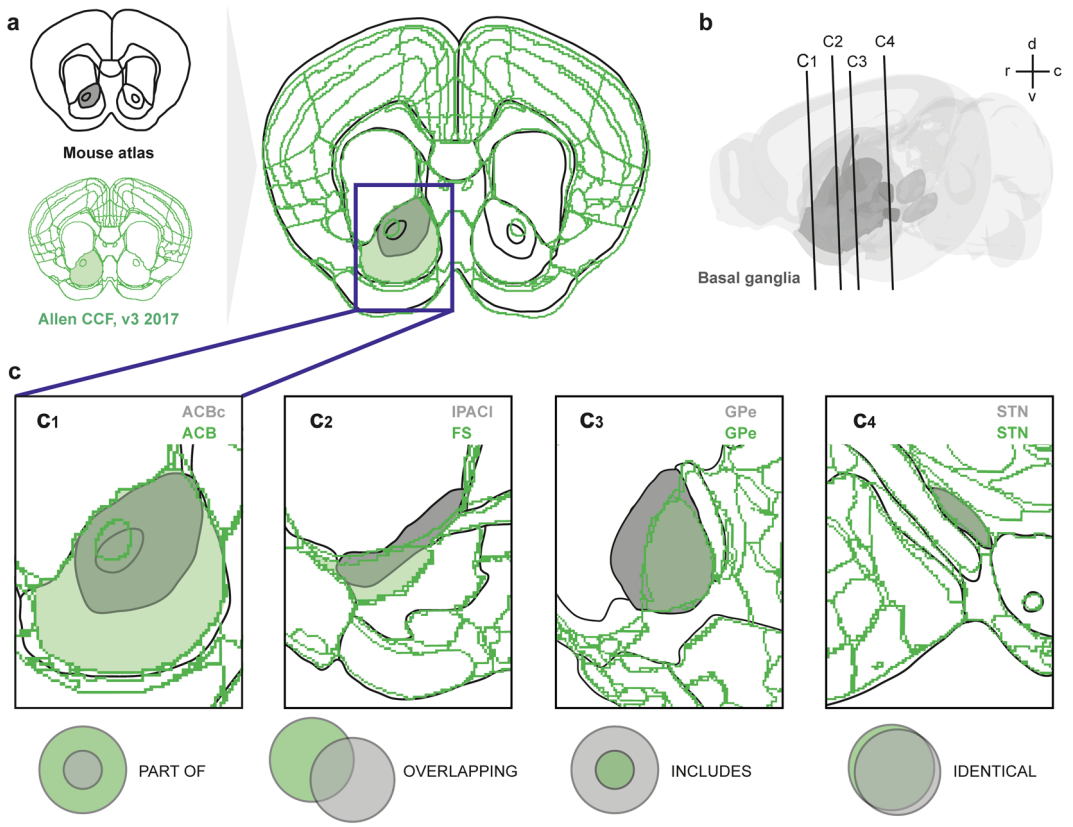


Fig. 3 Defining topological relationships of corresponding regions in different atlases. (**a**) Comparison of 2D coronal plates taken at the level of the genu of the corpus callosum in a fictive mouse atlas (black; drawn here for illustration purposes) to spatially matching custom plates through the 3D Allen Institute Common Coordinate Framework of the mouse brain (CCF, green[37];). By superimposing reference atlas plates with custom plates from the 3D atlas, it becomes possible to compare boundaries of regions in the two atlases. (**b**) The location of the plates shown in (**a**) and (**c**) is indicated in a transparent 3D rendering of the CCF atlas made using the Scalable Brain Atlas (https://scalablebrainatlas.incf.org/composer/?template=ABA_v3), with the basal ganglia shown in dark grey. After comparison of the co-registered atlas plates, the relationship between atlas regions is categorised as one of the types illustrated in (**a**), in which regions from the fictive atlas are shown in grey and CCF regions in green. A region is part of another region (c1), if its area is fully contained within the area of another region. For example, the fictive atlas has a region "nucleus accumbens core", which is completely part of the larger nucleus accumbens in the CCF. Overlapping regions (c2), pertains to the situation that corresponding regions in the two compared atlases partly occupy the same space, and partly not, as seen for the lateral interstitial nucleus of the posterior limb of the anterior commissure (IPACl) in the fictive atlas, which in some parts overlaps with the fundus of striatum (FS) and caudoputamen as defined in the CCF. A region includes another region (c3), if its area fully contains the area of the other region (c3), exemplified here by two versions of the external globus pallidus (GPe), which is larger in the fictive atlas than in the CCF. Identical regions are largely similar, with little or no areas of non-overlap (c4), exemplified here with the subthalamic nucleus (STN). Relationships are exemplified here for single sections, but were determined by comparison of co-registered atlas diagrams across entire regions. Abbreviations: ACBC, nucleus accumbens, core region; ACB, nucleus accumbens; FS, fundus of striatum; GPe, globus pallidus external segment; IPACl, lateral interstitial nucleus of the posterior limb of the anterior commissure; STN, subthalamic nucleus.

obtained by direct counts without any corrections were corrected using Abercrombie's formula[4,66] prior to calculating volumetric density. These calculations are elaborated in Supplementary File 5. We did not standardise total number estimates before entering these to the database, but rather indicated whether counts were uni- or bilateral. When it was not clear whether estimates were uni- or bilateral, we contacted the corresponding author of the paper to clarify. If no clarification was obtained, this field was indicated as "Unknown". In some cases assumptions, interpretations and slight modifications were made to give data similar formats, and we followed specific rules to ensure consistency throughout the data entry process. Details about this can be found in Supplementary File 6.

Sharing the database through the EBRAINS Knowledge Graph. We exported .csv files from all the tables in the database. In addition, we made and exported query tables containing selected metadata elements from multiple tables for quantitative estimates, distributions, and cell morphologies. We also created a version

of the database specifically designed to input data, and an Excel sheet configured for converting data from any density unit to volumetric densities or bilateral counts to unilateral ones. All of these elements (.csv files, empty database version, and Excel conversion sheet) are shared under a single dataset through EBRAINS[25].

Data Records

The database created here, hereafter referred to as the "Murine basal ganglia database", is shared via EBRAINS[25] (https://ebrains.eu). It contains information from 375 experiments reported in 245 sources; from these, we extracted 1228 quantitative estimates (501 total number estimates and 727 density estimates), 50 neuronal morphologies, and 18 distribution records of basal ganglia cellular parameters. The content of the Murine basal ganglia database is summarised in Fig. 4.

The shared dataset includes .csv files for all tables in the Murine basal ganglia database as well as the original. accdb file; these files contain the full set of metadata collected during the creation of the database. In addition, we share .csv and .xlsx files for data extracted from the database. These files contain all the numerical, distribution and morphology data available from the Murine basal ganglia database, with selected metadata that we considered relevant for most users. Furthermore, we have established a workflow to allow other researchers to expand upon the knowledge contained in the current version (detailed in usage notes below). To support the use of this workflow, we share an empty version of the Murine basal ganglia database (.accdb) with a spreadsheet (.xlsx), through which researchers can collect and / or contribute more information.

Technical Validation

In the following, we first consider how the search strings and selection criteria have affected the results of the PubMed search and content of the database. We then evaluate the validity of the graph data extraction procedure. Lastly, we assess and discuss the variability of a selection of the data contained in our database by summarising the information available about the number of tyrosine hydroxylase (TH) positive neurons in the substantia nigra, and the total number of neurons in the caudoputamen.

Selected papers and repositories. The most common reason for excluding papers were that they did not contain data of interest (54% of papers excluded) or that data were from experimentally manipulated animals without inclusion of a normal control group (15% of papers excluded). Among the studies in which relevant quantitative data had been obtained, 40–45% were excluded in each iteration of the search (11% of all papers) because data were not possible to normalise, due to lack of metadata necessary for comparing the data across studies or re-using them in a different context. Examples include papers where numbers were expressed per section or as percentage of control, or in rare cases, without specification of the unit of measurement. 8% of all search results were excluded because data were presented in graphs only, in each search this concerned 48–59% of the papers of interest with data that otherwise could have been normalised to a common unit of measurement. In a limited selection of papers presenting some data in text and other data in graphs we converted graph data to numeric data (see, Methods) to increase the amount of data extracted, but as this was time consuming it was not feasible to perform on a larger collection of data. In the end, 6% of papers were included. The percentages described here are based on data from the second and third search; the proportions of papers excluded based on the various criteria were relatively similar for the first search, except that the included percentage (3%) was lower since only completely untreated adult animals were included.

Searching and screening papers manually is a time consuming task, and in our literature search led to exclusion of more than 90% of papers. We observe that other literature mining projects have presented similar exclusion rates[67]. This illustrates that designing search strings that are both sensitive and specific is a significant challenge.

Validation of data extracted from graphs. Papers from the first iteration of the search that presented the same numbers both in graphs and text were used to validate the graph extraction approach. For these cases, we extracted the numbers and error measurements using the graph plot digitiser (see Methods), and compared the resulting numbers with those presented in the text. This showed a negligible discrepancy between means extracted from text and graph (0.08–1% difference), and relatively low differences between extracted standard errors (5–12% difference).

Variability in a selection of quantitative estimates from the murine basal ganglia. We here present summary data from some of the parameters available in the Murine basal ganglia database. To assess whether variance could be considerably reduced by selecting data obtained by certain methods, we sequentially filtered the data according to methodological metadata (see sections below for details).

Tyrosine hydroxylase positive neurons in the substantia nigra. The principal neurons of the substantia nigra are dopaminergic neurons, which can be visualised by using antibodies against the enzyme tyrosine hydroxylase. TH neurons contribute to motor behaviour by their projections to the striatum, and are frequently investigated in murine models for mechanisms of Parkinson's disease[26].

In our database, unilateral estimates of the total number of TH neurons in the substantia nigra, pars compacta of the adult (P56 and older) C57BL/6 mouse range from 1090 to 16145 (mean = 6065, SD = 3456, n = 30 estimates). The same range and very similar variation is seen when selecting only stereological studies (range = 1090 to 16145, mean = 6495, SD = 3503, n = 26 estimates). Further filtering of stereological estimates by excluding those that are anatomically non-specific or only partly covering the pars compacta, does not reduce variation either (range = 3360 to 16145, mean = 7706, SD = 3680, n = 14 estimates). Only two of the 30 total number estimates for the C57BL/6 mouse substantia nigra, pars compacta are connected to an antibody with a unique RRID; filtering results based on the exact primary antibody used is therefore not possible. For the adult (P60 and older)

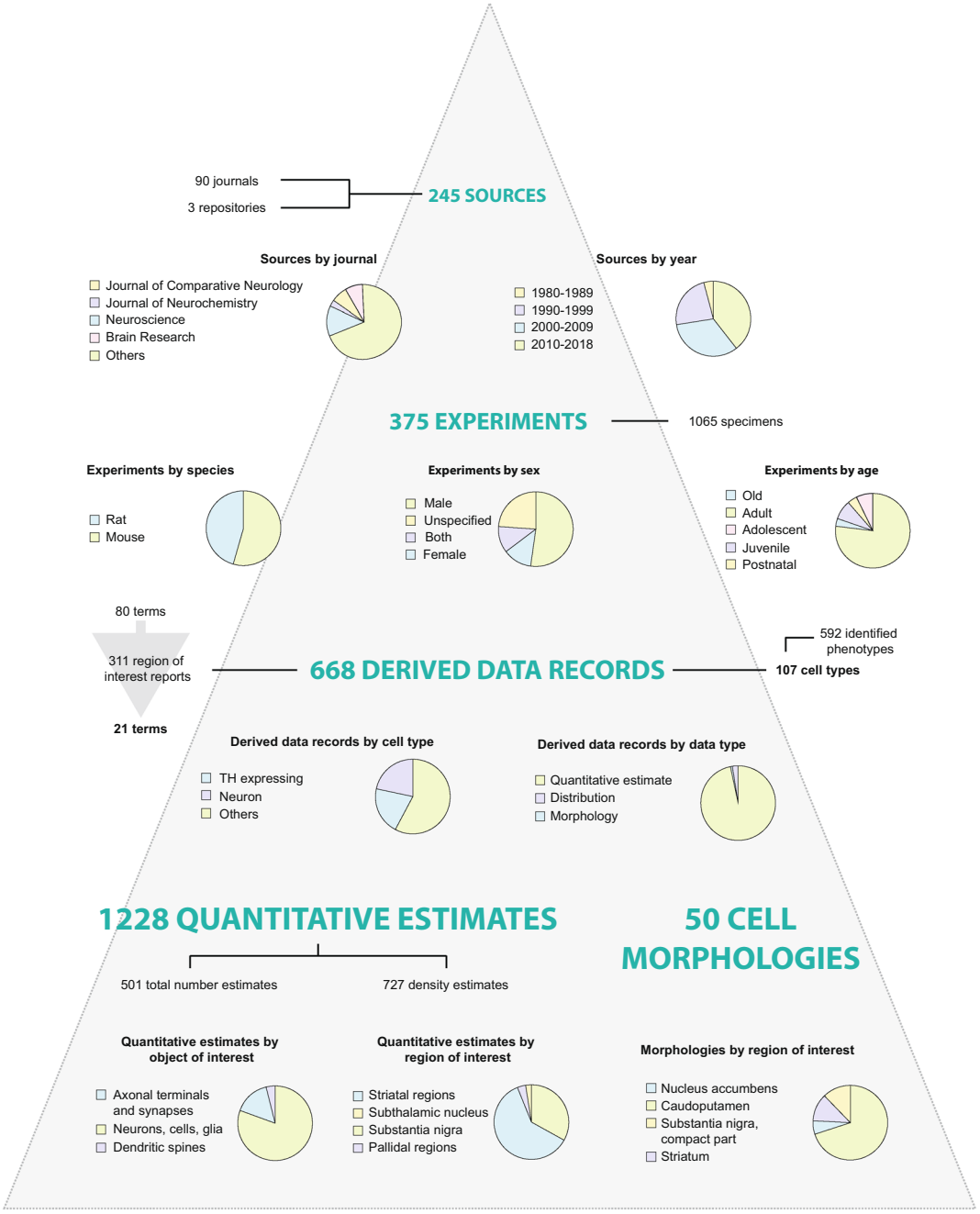


Fig. 4 Summary of database content. The current version of the database contains 1228 quantitative estimates and 50 cell morphologies, obtained from 375 experiments reported in 245 sources. Contents are sorted and displayed in pie charts according to key metadata elements. The most common journal sources include the Journal of Comparative Neurology, Journal of Neurochemistry, Neuroscience, and Brain Research. Most sources were published in 2000 or later. Experiments reported are primarily from male, adult animals, with slightly more mouse data included compared to rat data. Through our translation process for anatomical location metadata, 80 anatomical terms were translated to 21 terms found in standard rat and mouse atlases. Data on a total of 100 cell types are included in the database, with approximately half of the 668 derived data records related to TH positive neurons or all neurons. Of the 1228 quantitative estimates found in the database, 501 are total number estimates and 727 are density estimates. These are mainly estimated numbers of neurons, cells or glia, with substantia nigra and striatal regions heavily represented. The 50 cell morphologies in the database are mainly from striatal regions (nucleus accumbens, caudoputamen, or striatum overall). Additional metadata are represented in the database, which is available from the EBRAINS Knowledge Graph[25].

rat (all strains), the range of unilateral values in the database is 3260 to 11969 (mean = 7733, SD = 3252, n = 8 estimates). Box plots summarizing these estimates and similar ones from the whole substantia nigra are given in Fig. 5a.

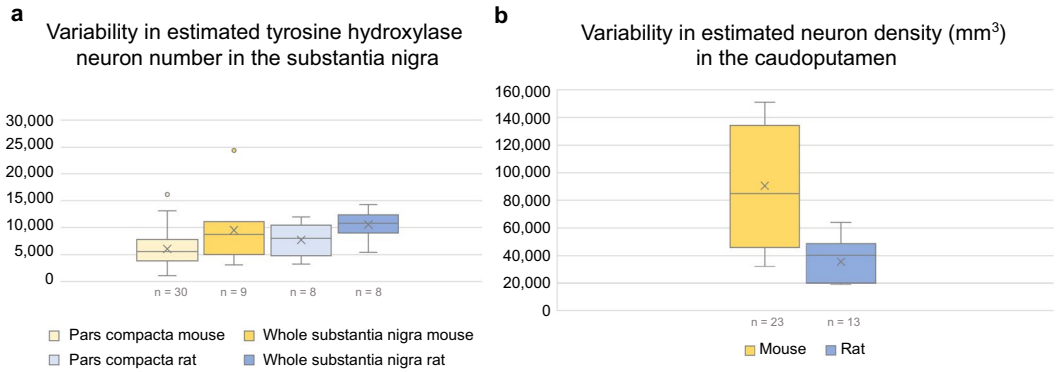


Fig. 5 Variability of estimates from the literature. Figure showing summary data for estimates from the database. The whiskers represents the values that fall within 1.5 times the interquartile range. (**a**) Box plot showing total number estimates of TH positive cells in one hemisphere of the substantia nigra pars compacta (light yellow and light blue boxes for mouse and rat, respectively) and the whole substantia nigra (including the compact, reticular and lateral parts, which in some studies also may have included parts of the ventral tegmental area; dark yellow and dark blue boxes for mouse and rat, respectively). (**b**) Box plot showing neuron density estimates in the caudoputamen of mice (yellow box) and rat (blue box). The *n* represents the number of estimates, but note that more than one of these could originate from the same publication or repository source.

Neuron numbers and densities in the caudoputamen. The caudate-putamen complex (hereafter referred to as the caudoputamen) is the largest part of the basal ganglia, receiving axonal projections from the cerebral cortex, and extending projections to several other parts of the basal ganglia circuitry[68]. There are two main types of principal neurons in the caudoputamen, identifiable by the different types of dopamine receptors they possess[26]. Because well-validated and replicated antibodies against these receptors are lacking[69], studies of the caudoputamen frequently assess total neuron numbers using histochemistry or neuronal markers such as NeuN antibodies.

Unilateral estimates in the database representing the total number of neurons in the caudoputamen of adult mice (all strains) range from 856649 to 1711615 (mean = 1107325, SD = 296707, n = 10 estimates). Note that six of these estimates come from the same study. Estimates of neuron density range from 32166 to 151112 per cubic millimetre (mean = 90407, SD = 42133, n = 23 estimates). The range of density estimates is the same and variability not reduced by selecting stereological estimates only (mean = 88705, SD = 44009, n = 18 studies). In rats, only very few estimates of total numbers for the caudoputamen are available in the database. The estimated neuron density in the adult rat caudoputamen varies from 19129 to 64050 neurons per cubic millimetre (mean = 35529, SD = 15029, n = 13 estimates). The neuron density estimates for caudoputamen are summarised in box plots in Fig. 5b.

Possible reasons for observed variability. Our assessment of the variability of quantitative neuroanatomical data from the substantia nigra show that for TH expressing neurons in the pars compacta of C57BL/6 mice, the reported numbers range from approximately 1000 to over 16,000 cells unilaterally. High variability is also seen in the caudoputamen data. Interestingly, the reported neuron density (per cubic millimetre) is on average ~twice as high in the mouse than in the rat. Although the numbers reported within the species varies a lot, the ratio between the mean densities correspond well with estimated scaling rules between rat and mouse brains[70]. For the mouse caudoputamen, estimates of total neuron numbers in the database range from 856649 to 1711615 in one hemisphere. Few studies include the range of values collected in addition to summary statistics, but it is clear that the variability between studies is much higher than that within studies. For example, in a study comparing neuron numbers in the caudoputamen across different mouse strains[71], the difference between the bilateral average of the groups with the highest and lowest number was 324926. It is thus highly unlikely that the range we observe across studies, of almost one million cells unilaterally, can be attributed solely to biological variance. Instead, the reasons for the large variation in numbers reported from within a region are likely to be manifold. Due to the wide-spread lack of methodological metadata in papers, the size of groups containing estimates obtained by clearly defined and similar methods was too small to support formal statistical analysis on differences in variability. Collection of more data to the Murine basal ganglia database, combined with improved reporting practices, could allow such analyses in the future. Nevertheless, we believe that the present data collection, shared through the EBRAINS Knowledge Graph, can be useful for finding and comparing published data. The ability to filter the data based on metadata elements might also be useful to select appropriate data, depending on the need of the user. Combined with our defined workflows for contributing more information, we believe these results will make it easier to select, organise, compare and share quantitative information from the literature or from new analyses in the future. We describe these uses of the database in detail below.

Usage Notes

Our database is shared through the EBRAINS Knowledge Graph as part of a dataset entitled "Database of quantitative cellular and subcellular morphological properties from rat and mouse basal ganglia"[25]. It comprises three main parts (see Data records for details): 1) the Murine basal ganglia database (Database_v1.accdb); 2) spreadsheets with all the quantitative estimates, morphologies and distributions contained in the Murine basal ganglia

database (files in .xlsx and .csv format) with selected metadata; and 3) an empty version of the Murine basal ganglia database (Input_database.accdb) and a spreadsheet (Input_sheet .xlsx) facilitating collection of new data. We here briefly explain how researchers with different interests may utilise different parts of this dataset. These descriptions are intended as example use cases, and the reader is referred to other sources[72,73] for guides on the use of Microsoft Access and Excel (RRID:SCR_016137). Because maintaining and updating information is a challenge with any database that is seldom addressed, we go on to describe a workflow through which other researchers can organise and share more data, using shared database and spreadsheet templates.

Using the Murine basal ganglia database and the data extracted from it. *Exploring the reported numbers of tyrosine hydroxylase neurons in substantia nigra.* A researcher wants to look up the reported numbers of TH neurons in substantia nigra. Having downloaded the dataset titled "Database of quantitative cellular and subcellular morphological properties from rat and mouse basal ganglia"[25], (s)he opens the README file to get a quick overview of the contents. There, (s)he sees that the Data Extracts-folder contains queries that include all the numbers available in the database. (S)he finds that such extracts are likely to meet his/her questions, and navigates to the relevant folder. S(h)e opens the.xlxs file called "Cell counts" and selects the sheet called "Total number estimates". The first three columns show the cell types that have been quantified, the species, and the regions of interest. The researcher filters the records to "Tyrosine hydroxylase expressing" cells, "Mus musculus" and "Pars compacta" (to simultaneously filter multiple columns in Microsoft Excel, select all the columns to be filtered, and under the Data tab click "Filter"). This yields 70 records, each one with accompanying metadata elements related to the animals, counting method, and region of interest. To explore the data further, e.g. by extracting descriptive measurements, (s)he copies the filtered records to a new sheet (in Microsoft Excel, go to Find & Select, click "Go to special" and select "Visible cells only").

Finding studies using a specific primary antibody. A researcher has used immunohistochemistry to visualize parvalbumin positive neurons, and quantified labelled cells using stereological analysis. To verify the results (s)he is now interested in finding quantitative data from studies where the same antibody has been used. (S)he downloads the dataset titled "Database of quantitative cellular and subcellular morphological properties from rat and mouse basal ganglia"[25] from EBRAINS, and upon looking at the "Cell counts" data extracts finds that they do not contain metadata specifying the antibody used. (S)he therefore navigates to the Database-folder and opens the.accdb file and the text file called "Tables_description". In the text file, (s)he finds that antibodies are stored in the lookup table "Reporters" with connections to the table "Sources" via other tables. (S)he navigates to the Create table and clicks the Query Wizard. After selecting the Simple Query Wizard, (s)he selects the "Reporter name" and "Reporter unique ID" fields from the "Reporters" table and the "Source name" and "Source ID" field from the "Sources" table. (S)he clicks "Finish" and is presented with a query including a list of antibodies, their unique RRIDs, and the name and DOI of the source in which they were used. (S)he clicks the "Reporter unique ID" column header and scrolls to see if the antibody of interest (RRID:AB_10000344) is among the listed IDs. It is, and (s)he filters the list to these records. This yields a list of four studies where the antibody has been used. The researcher can now look for results from these studies in the Cell counts data extract by filtering it to the relevant Source names, or look up the original papers by use of the DOIs.

Overviewing the methodological parameters of a study. Upon identifying the studies using an antibody of interest, the researcher described above wants to get a quick overview of the methods used in these studies. In the Tables description-file, (s)he reads that the tables "Specimens" and "Specimen_treatments" contain information related to the treatment of tissue reported in included papers. (S)he therefore creates a new query, including the "Source name" from the "Sources" table, the "Specimen form" from the "Specimens" table and the fields "Solution", "Purpose", "Time", "Time unit" and "Temperature" from the "Specimen treatments" table. Since (s)he is primarily interested in the methods related to parvalbumin stained material, (s)he also includes the "Cell type putative" field from the table "Derived data records". She clicks "Finish", and filters the resulting query to his/her studies of interest by clicking the "Source name" column header and selecting the relevant Source names. (S)he also clicks the "Cell type putative" column header and uses the text filter to select only records that contains "Parvalbumin". This yields 13 records summarizing the treatments used for each specimen.

Using the workflow for harvesting, organizing and updating neuroanatomical data. In order to compare numbers reported across studies, it is first necessary to systematically extract relevant data and metadata and to standardise these to common units and concepts. We here present a workflow enabling users to harvest and organise their quantitative neuroanatomical data from the literature or public databases. This workflow includes a template version of the Murine basal ganglia database (with forms supporting input of largely standardised metadata) available through the dataset hosted in the EBRAINS Knowledge Graph[25], a novel procedure for translating terms for regions of interest to common terminology, and a preliminary approach for mapping data to cell-types of interest. We include steps through which other researchers can enter new information to the database and share this with the broader community. The workflow is summarised in Fig. 6.

Translation of terms across neuroanatomical nomenclatures. A key part of the overall workflow is the translation of semantic terms existing in different reference atlases to terms used in the standard reference atlases used by EBRAINS (Supplementary Fig. 1). These include the Waxholm Space atlas of the Rat Brain[34,35,74] and the Allen Mouse Brain Common Coordinate Framework[37]. Anatomical metadata found in sources essentially enters one of three routes. In the first route, terms that are consistent with the nomenclature for one of the EBRAINS standard atlases are directly entered in the database. In the second route, terms that are consistent with another atlas nomenclature are translated to the closest matching region term in the relevant EBRAINS atlas. The basis for

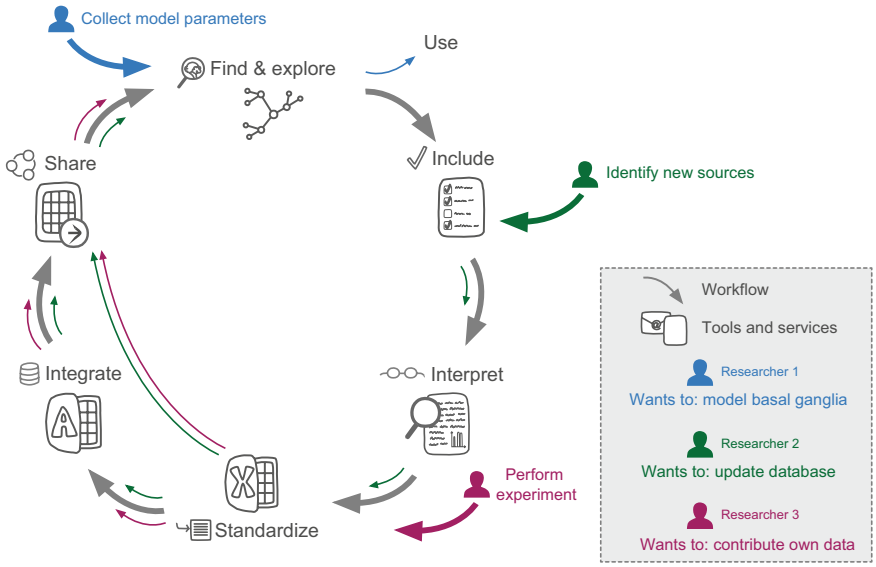


Fig. 6 Workflow for integration of derived data. Schematic representation of the main workflow described in this paper, as well as the tools and services used (in grey). Coloured arrows indicate specific ways different users may interact with the database or the workflow, today or in the future. Blue arrows illustrate how a researcher may use the database to identify data to constrain a computational model of the basal ganglia. Through the EBRAINS Knowledge Graph, derived data may be found as a shared dataset[25], downloaded and explored. Green and purple arrows show how researchers may use the workflow and resources developed to update the database in the future, identifying and interpreting new sources (green arrows) or by contributing own data (purple arrows). In the workflow, data are included by use of specific criteria, and interpreted manually. Derived data are extracted and standardised to common units of measurement using a custom Excel workbook. The Excel workbook is designed for input of data as well as minimum metadata according to the EBRAINS standard. Optionally, more extensive metadata may be integrated through a Microsoft Access database input portal. Once organised, derived data may be shared through the EBRAINS curation support, through which it can be made available as part of the collection found in the Knowledge Graph.

making such a translation is given by the spatial relationships between regions delineated in the different atlases used, and regions in the EBRAINS standard atlases (see Methods for details); these relationships are available as datasets through the EBRAINS Knowledge Graph[61,62]. The third route for inserting anatomical metadata found in sources, are for terms that are not consistent with any standard nomenclature. These are treated as "custom" terms, and translated to the closest EBRAINS atlas term by using the documentation available from the source. This was the most commonly used route in the current project, used for ~62% of reported regions. Note that 4% of the sources used an atlas for which relationships to EBRAINS atlases were not available, and these also entered the custom region translation route.

Updating the Murine basal ganglia database. A database is only up-to-date as long as the data are maintained and expanded with new information. In addition to sharing the content aggregated and organised through this project, we therefore outline how researchers could contribute to the Murine basal ganglia database in the future. Researchers might want to add more data from the literature (green arrows, Fig. 6) or from own experiments (purple arrows, Fig. 6). The first step for anyone wishing to add more data from the published literature is to identify potential new sources. This could be done through a literature search similar to that described in this paper (with date filters constraining the search to the period after the current search was performed, see Supplementary File 2 for a list of the search strings used here). Alternatively, advances in text mining might yield opportunities for more automatic search strategies[75]. The next step in the workflow is interpretation: the source needs to be examined manually to identify the relevant data and metadata elements to be extracted. Once produced or identified through the literature, data should be extracted and metadata standardised. For this purpose, we share an Excel workbook with sheets where data can be entered in any format. Upon insertion of the number and unit, calculated fields standardise data to represent number per square or cubic micro- or millimetre. For cell counts, volumetric densities are also estimated from 2D counts given that section thickness is provided, according to the calculations described in Supplementary File 5. The Excel workbook is also tailored for input of basic metadata as required by EBRAINS (https://ebrains.eu/). It may be used in a relatively simple route to collecting data and contributing to the database (long green and purple lines in Fig. 6). Alternatively, it may be used as a means to organise and standardise data before entering it with extended metadata in the database. For this purpose, we share an empty version of the Murine basal ganglia database (an "input portal") specifically designed to add new data with the full extent of metadata collected for the current project. The Excel sheet and Access database tailored for input are shared as .xlxs and .accdb files, respectively, and can be downloaded together with the Murine basal ganglia database through the dataset hosted at EBRAINS Knowledge Graph[25]. In our experience,

the interpretation, extraction, standardisation and integration steps might require from half an hour to several hours per publication; generally, less time is required to integrate data that is provided in tabular format and using standard units of measurement (square or cubic micro- or millimetres), with a clearly described methodology. The files shared here to facilitate collection of new data may be populated and stored locally by the user, or shared with the community. The last step of the workflow outlined here (Fig. 6) is thus sharing the data. This could be done through any data sharing platform, e.g. Zenodo (www.zenodo.org) or Figshare (www.figshare.com). The EBRAINS curation service (curation-support@ebrains.eu) and Knowledge Graph offers the advantage of being tailored to neuroscience data, and would allow for new data collections to be linked to the version of the Murine basal ganglia database presented here[25].

Code availability

The QuickNII (RRID:SCR_016854) tool was used for spatial co-registration of atlases. Microsoft Access 2016 was used to create the database.

Received: 20 January 2020; Accepted: 4 June 2020;
Published online: 06 July 2020

References

1. Egger, R., Dercksen, V., Udvary, D., Hege, H.-C. & Oberlaender, M. Generation of dense statistical connectomes from sparse morphological data. *Front. Neuroanat.* **8**, 1–18 (2014).
2. Markram, H. *et al*. Reconstruction and simulation of neocortical microcircuitry. *Cell* **163**, 456–492 (2015).
3. Bezaire, M. J. & Soltesz, I. Quantitative assessment of CA1 local circuits: knowledge base for interneuron-pyramidal cell connectivity. *Hippocampus* **23**, 751–785 (2013).
4. Abercrombie, M. Estimation of nuclear population from microtome sections. *Anat. Rec.* **94**, 239–247 (1946).
5. Schmitz, C. & Hof, P. Design-based stereology in neuroscience. *Neuroscience* **130**, 813–831 (2005).
6. Brændgaard, H. & Gundersen, H. J. G. The impact of recent stereological advances on quantitative studies of the nervous system. *J. Neurosci. Methods* **18**, 39–78 (1986).
7. Bjaalie, J., Diggle, P., Nikundiwe, A., Karagulle, T. & Brodal, P. Spatial segregation between populations of ponto-cerebellar neurons: Statistical analysis of multivariate spatial interactions. *Anat. Rec.* **231**, 510–523 (1991).
8. Prodanov, D., Nagelkerke, N. & Marani, E. Spatial clustering analysis in neuroanatomy: Applications of different approaches to motor nerve fiber distribution. *J. Neurosci. Methods* **160**, 93–108 (2007).
9. West, M. J., Østergaard, K., Andreassen, O. A. & Finsen, B. Estimation of the number of somatostatin neurons in the striatum: An *in situ* hybridization study using the optical fractionator method. *J. Comp. Neurol.* **370**, 11–22 (1996).
10. Oorschot, D. Total number of neurons in the neostriatal, pallidal, subthalamic, and substantia nigral nuclei of the rat basal ganglia: A stereological study using the cavalieri and optical disector methods. *J. Comp. Neurol.* **599**, 580–599 (1996).
11. Yu, Z. *et al*. Nitrated α-synuclein induces the loss of dopaminergic neurons in the substantia nigra of rats. *PLoS One* **5**, (2010).
12. Singh, A. *et al*. Long term exposure to cypermethrin induces nigrostriatal dopaminergic neurodegeneration in adult rats: postnatal exposure enhances the susceptibility during adulthood. *Neurobiol. Aging* **33**, 404–415 (2012).
13. Bornmann, L. & Mutz, R. Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *J. Assoc. Inf. Sci. Technol.* **66**, 2215–2222 (2015).
14. van Strien, N. M., Cappaert, N. L. & Witter, M. P. The anatomy of memory: an interactive overview of the parahippocampal-hippocampal network. *Nat Rev Neurosci* **10**, 272–282 (2009).
15. Bota, M., Dong, H.-W. & Swanson, L. Brain Architecture Management System. *Neuroinformatics* **3**, 015–048 (2005).
16. Wheeler, D. *et al*. Hippocampome.org: a knowledge base of neuron types in the rodent hippocampus. *Elife* **4**, 1138–1142 (2015).
17. Ascoli, G., Donohue, D. & Halavi, M. NeuroMorpho.Org: A central resource for neuronal morphologies. *J. Neurosci.* **27**, 9247–9251 (2007).
18. Erö, C., Gewaltig, M., Keller, D. & Markram, H. A cell atlas for the mouse brain. *Front. Neuroinform.* **12**, 1–16 (2018).
19. Kim, Y. *et al*. Brain-wide maps reveal stereotyped cell-type-based cortical architecture and subcortical sexual dimorphism. *Cell* **171**, 456–469.e22 (2017).
20. Murakami, T. *et al*. A three-dimensional single-cell-resolution whole-brain atlas using CUBIC-X expansion microscopy and tissue clearing. *Nat. Neurosci.* **21**, 625 (2018).
21. Obeso, J. *et al*. The basal ganglia in Parkinson's disease: Current concepts and unexplained observations. *Ann. Neurol.* **64**, S30–S46 (2009).
22. Bunner, K. D. & Rebec, G. V. Corticostriatal dysfunction in Huntington's disease: The basics. *Front. Hum. Neurosci.* **10**, 317 (2016).
23. Vidyadhara, D. J., Yarreiphang, H., Raju, T. R. & Alladi, P. A. Admixing of MPTP-resistant and susceptible mice strains augments nigrostriatal neuronal correlates to resist MPTP-induced neurodegeneration. *Mol. Neurobiol.* **54**, 6148–6162 (2017).
24. Baquet, Z., Williams, D., Brody, J. & Smeyne, R. A comparison of model-based (2D) and design-based (3D) stereological methods for estimating cell number in the substantia nigra pars compacta (SNpc) of the C57BL/6J mouse. *Neuroscience* **161**, 1082–1090 (2009).
25. Bjerke, I., Puchades, M., Bjaalie, J. G. & Leergaard, T. Database of quantitative cellular and subcellular morphological properties from rat and mouse basal ganglia. *Human Brain Project Neuroinformatics Platform* https://doi.org/10.25493/DYXZ-76U (2019).
26. Gerfen, C. R. & Bolam, J. P. The neuroanatomical organization of the basal ganglia. *Handb. Behav. Neurosci.* **24**, 3–32 (2016).
27. Olmos, J. & Heimer, L. The concepts of the ventral striatopallidal system and extended amygdala. *Ann. N. Y. Acad. Sci.* **877**, 1–32 (1999).
28. Gupta, A. *et al*. Federated access to heterogeneous information resources in the neuroscience information framework (NIF). *Neuroinformatics* **6**, 205–217 (2008).
29. Polavaram, S., Gillette, T., Parekh, R. & Ascoli, G. Statistical analysis and data mining of digital reconstructions of dendritic morphologies. *Front. Neuroanat.* **8**, 1–16 (2014).
30. Yates, S. & Puchades, M. Extraction of parvalbumin positive cells from an Allen mouse brain *in situ* hybridisation experiment. *Human Brain Project Neuroinformatics Platform* https://doi.org/10.25493/6DYS-M3W (2019).
31. Martone, M. *et al*. The Cell Centered Database project: An update on building community resources for managing and sharing 3D imaging data. *J. Struct. Biol.* **161**, 220–231 (2008).
32. Paxinos, G. & Watson, C. *The rat brain in stereotaxic coordinates*. (Elsevier Inc (2013).
33. Swanson, L. *Brain Maps III: Structure of the rat brain*. (Elsevier (2004).
34. Papp, E., Leergaard, T. B., Calabrese, E., Johnson, G. A. & Bjaalie, J. G. Waxholm Space atlas of the Sprague Dawley rat brain. *Neuroimage* **97**, 374–386 (2014).

35. Kjonigsen, L., Lillehaug, S., Bjaalie, J., Witter, M. & Leergaard, T. Waxholm Space atlas of the rat brain hippocampal region: Three-dimensional delineations based on magnetic resonance and diffusion tensor imaging. *Neuroimage* **108**, 441–449 (2015).
36. Paxinos, G. & Franklin, K. *The mouse brain in stereotaxic coordinates*. (Academic Press (2012).
37. Oh, S. *et al*. A mesoscale connectome of the mouse brain. *Nature* **508**, 207–214 (2014).
38. Papp, E. A., Leergaard, T. B., Calabrese, E., Allan Johnson, G. & Bjaalie, J. G. Addendum to "Waxholm Space atlas of the Sprague Dawley rat brain" [NeuroImage 97 (2014) 374-386]. *Neuroimage* **105**, 561–562 (2015).
39. Puchades, M., Csucs, G., Ledergerber, D., Leergaard, T. & Bjaalie, J. Spatial registration of serial microscopic brain images to three-dimensional reference atlases with the QuickNII tool. *PLoS One* **14**, (2019).
40. Franklin, K. & Paxinos, G. *The mouse brain in stereotaxic coordinates*. (Academic Press (1996).
41. Franklin, K. & Paxinos, G. *The mouse brain in stereotaxic coordinates*. (Academic Press (2007).
42. Paxinos, G. & Watson, C. *The rat brain in stereotaxic coordinates*. (Academic Press (1998).
43. Paxinos, G. & Watson, C. *The rat brain in stereotaxic coordinates*. (Academic Press (2007).
44. Paxinos, G. & Watson, C. *The rat brain in stereotaxic coordinates*. (Academic Press (1986).
45. Paxinos, G. & Watson, C. *The rat brain in stereotaxic coordinates*. (Elsevier (2005).
46. Swanson, L. *Brain Maps: Structure of the rat brain*. (Elsevier (1992).
47. Swanson, L. *Brain Maps II: Structure of the rat brain*. (Elsevier (1998).
48. Bjerke, I., Schlegel, U., Puchades, M., Bjaalie, J. & Leergaard, T. Paxinos & Watson's "The Rat Brain in Stereotaxic Coordinates" (3rd edition) spatially registered to the Waxholm Space atlas of the rat brain. *Human Brain Project Neuroinformatics Platform* https://doi.org/10.25493/KNB2-GMN (2019).
49. Bjerke, I., Schlegel, U., Puchades, M., Bjaalie, J. & Leergaard, T. Paxinos & Watson's "The Rat Brain in Stereotaxic Coordinates" (5th edition) spatially registered to the Waxholm Space atlas of the rat brain. *Human Brain Project Neuroinformatics Platform* https://doi.org/10.25493/KQ5K-S0D (2019).
50. Bjerke, I., Schlegel, U., Puchades, M., Bjaalie, J. & Leergaard, T. Swanson's "Brain Maps: Structure of the Rat Brain" (3rd edition) spatially registered to the Waxholm Space atlas of the rat brain. *Human Brain Project Neuroinformatics Platform* https://doi.org/10.25493/ZFXB-23F (2019).
51. Bjerke, I., Schlegel, U., Puchades, M., Bjaalie, J. & Leergaard, T. Paxinos & Watson's "The Rat Brain in Stereotaxic Coordinates" (4th edition) spatially registered to the Waxholm Space atlas of the rat brain. *Human Brain Project Neuroinformatics Platform* https://doi.org/10.25493/W3R1-R4A (2019).
52. Bjerke, I., Schlegel, U., Puchades, M., Bjaalie, J. & Leergaard, T. Paxinos & Watson's "The Rat Brain in Stereotaxic Coordinates" (1st edition) spatially registered to the Waxholm Space atlas of the rat brain. *Human Brain Project Neuroinformatics Platform* https://doi.org/10.25493/YRKH-626 (2019).
53. Bjerke, I., Schlegel, U., Puchades, M., Bjaalie, J. & Leergaard, T. Swanson's "Brain Maps: Structure of the Rat Brain" (4th edition) spatially registered to the Waxholm Space atlas of the rat brain. *Human Brain Project Neuroinformatics Platform* https://doi.org/10.25493/486N-966 (2019).
54. Bjerke, I., Schlegel, U., Puchades, M., Bjaalie, J. & Leergaard, T. Paxinos & Watson's "The Rat Brain in Stereotaxic Coordinates" (7th edition) spatially registered to the Waxholm Space atlas of the rat brain. *Human Brain Project Neuroinformatics Platform* https://doi.org/10.25493/APWV-37H (2019).
55. Bjerke, I., Schlegel, U., Puchades, M., Bjaalie, J. & Leergaard, T. Swanson's "Brain Maps: Structure of the Rat Brain" (2nd edition) spatially registered to the Waxholm Space atlas of the rat brain. *Human Brain Project Neuroinformatics Platform* https://doi.org/10.25493/EEQA-9RM (2019).
56. Bjerke, I., Schlegel, U., Puchades, M., Bjaalie, J. & Leergaard, T. Swanson's "Brain Maps: Structure of the Rat Brain" (1st edition) spatially registered to the Waxholm Space atlas of the rat brain. *Human Brain Project Neuroinformatics Platform* https://doi.org/10.25493/ZB03-H5G (2019).
57. Bjerke, I., Schlegel, U., Puchades, M., Bjaalie, J. & Leergaard, T. Franklin & Paxinos' "The Mouse Brain in Stereotaxic Coordinates" (3rd edition) spatially registered to the Allen mouse brain Common Coordinate Framework. *Human Brain Project Neuroinformatics Platform* https://doi.org/10.25493/1BT9-YYD (2019).
58. Bjerke, I., Schlegel, U., Puchades, M., Bjaalie, J. & Leergaard, T. Paxinos & Watson's "The Rat Brain in Stereotaxic Coordinates" (6th edition) spatially registered to the Waxholm Space atlas of the rat brain. *Human Brain Project Neuroinformatics Platform* https://doi.org/10.25493/XQ8J-TNE (2019).
59. Bjerke, I., Schlegel, U., Puchades, M., Bjaalie, J. & Leergaard, T. Franklin & Paxinos' "The Mouse Brain in Stereotaxic Coordinates" (4th edition) spatially registered to the Allen Mouse Common Coordinate Framework. *Human Brain Project Neuroinformatics Platform* https://doi.org/10.25493/WFCZ-FSN (2019).
60. Bjerke, I., Schlegel, U., Puchades, M., Bjaalie, J. & Leergaard, T. Paxinos & Franklin's "The Mouse Brain in Stereotaxic Coordinates" (2nd edition) spatially registered to the Allen Mouse Common Coordinate Framework. *Human Brain Project Neuroinformatics Platform* https://doi.org/10.25493/BTKK-CRY (2019).
61. Bjerke, I., Puchades, M., Bjaalie, J. & Leergaard, T. Comparability of basal ganglia delineations across different mouse brain atlases. *Human Brain Project Neuroinformatics Platform* https://doi.org/10.25493/MWAS-3S6 (2019).
62. Bjerke, I., Puchades, M., Bjaalie, J. & Leergaard, T. Comparability of basal ganglia delineations across different rat brain atlases. *Human Brain Project Neuroinformatics Platform* https://doi.org/10.25493/D2M9-BSK (2019).
63. Hamilton, D. *et al*. Name-calling in the hippocampus (and beyond): coming to terms with neuron types and properties. *Brain Informatics* **4**, 1–12 (2017).
64. Ascoli, G. *et al*. Petilla terminology: nomenclature of features of GABAergic interneurons of the cerebral cortex. *Nat. Rev. Neurosci.* **9**, 557–568 (2008).
65. Keller, D., Erö, C. & Markram, H. Cell densities in the mouse brain: A systematic review. *Front. Neuroanat.* **12**, 83 (2018).
66. Coggeshall, R. A consideration of neural counting methods. *Trends Neurosci.* **15**, 9–13 (1992).
67. Sugar, J., Witter, M., van Strien, N. & Cappaert, N. The retrosplenial cortex: intrinsic connectivity and connections with the (para) hippocampal region in the rat. An interactive connectome. *Front. Neuroinform.* **5**, 1–13 (2011).
68. Voorn, P., Vanderschuren, L., Groenewegen, H., Robbins, T. & Pennartz, C. Putting a spin on the dorsal-ventral divide of the striatum. *Trends Neurosci.* **27**, 468–474 (2004).
69. Cullity, E., Madsen, H., Perry, C. & Kim, J. Postnatal developmental trajectory of dopamine receptor 1 and 2 expression in cortical and striatal brain regions. *J. Comp. Neurol.* 1–17 (2018).
70. Herculano-Houzel, S., Mota, B. & Lent, R. Cellular scaling rules for rodent brains. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 12138–12143 (2006).
71. Rosen, G. D. & Williams, R. W. Complex trait analysis of the mouse striatum: independent QTLs modulate volume and neuron number. *BMC Neurosci.* **2**, 5 (2001).
72. Barrows, A., Young, M. & Stockman, J. *Microsoft Access 2010 all-in-one for dummies*. (Wiley Publishing (2010).
73. Frye, C. *Microsoft Excel 2019*. (Microsoft Press (2019).
74. Osen, K., Imad, P., Wennberg, A., Papp, E. & Leergaard, T. Waxholm Space atlas of the rat brain auditory system: Three-dimensional delineations based on structural and diffusion tensor magnetic resonance imaging. *Neuroimage* **199**, 38–56 (2019).
75. French, L. *et al*. Text mining for neuroanatomy using WhiteText with an updated corpus and a new web application. *Front. Neuroinform.* **9**, 13 (2015).

Acknowledgements

We thank Krister Andersson, Camilla Blixhavn, Heidi Kleven, Martin Øvsthus, Lyuba Zehl and Oliver Schmid for useful discussions, and Ulrike Schlegel for help with sharing datasets through the EBRAINS Knowledge Graph. This work was funded by the European Union's Horizon 2020 Framework Programme for Research and Innovation Programme under the Specific Grant Agreement No. 785907 (Human Brain Project SGA2), Specific Grant Agreement No. 945539 (Human Brain Project SGA3), and The Research Council of Norway under Grant Agreement No. 269774 (INCF Norwegian Node). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions

I.E.B. contributed to conceiving of the study; collected, organised and analysed the data; and co-authored the manuscript. M.A.P. and J.G.B. contributed to conceiving and supervising the study. T.B.L. conceived and supervised the study, and co-authored the manuscript. All authors reviewed and approved of the manuscript.

Competing Interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41597-020-0550-3.

Correspondence and requests for materials should be addressed to T.B.L.

Reprints and permissions information is available at www.nature.com/reprints.

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

III

Brain-wide quantitative analysis reveals complementary patterns of parvalbumin and calbindin neurons in executive and modulatory regions

Ingvild E. Bjerke[1], Sharon C. Yates[1], Arthur Laja[2], Menno P. Witter[2], Maja A. Puchades[1], Jan G. Bjaalie[1] & Trygve B. Leergaard[1]*

[1] Department of Molecular Medicine, Institute of Basic Medical Sciences, University of Oslo, Norway

[2] Kavli Institute, Norwegian University of Science and Technology, Norway

* Correspondence: t.b.leergaard@medisin.uio.no

SUMMARY

The calcium-binding proteins parvalbumin and calbindin are expressed in neuronal populations regulating brain networks involved in spatial navigation, memory processes and social interactions. Information about their numbers across brain regions is required to understand the functional roles of parvalbumin- and calbindin-expressing neurons, and facilitate building of realistic computational models, but such data are scarcely available. We performed brain-wide analysis of immunohistochemically stained parvalbumin and calbindin sections and here show that parvalbumin and calbindin neurons distribute in largely complementary patterns across the mouse brain. Our data reveal that parvalbumin neurons dominate in areas related to sensorimotor processing and navigation, while calbindin neurons prevail in regions reflecting behavioural states. We also find that parvalbumin neurons distribute according to similar principles in the hippocampal region of the rat and mouse brain. Comparison of our results to previous reports shows that neuron numbers vary, while patterns of relative densities and numbers are consistent.

INTRODUCTION

Transient increases in intracellular calcium concentrations play a critical role in the regulation of neuronal excitability, neurotransmitter release and synaptic plasticity (Berridge, 1998). The spatial and temporal dynamics of such calcium signals can be modulated by calcium binding proteins, which are widely expressed in the nervous system (Schwaller, 2010). Two such calcium binding proteins, parvalbumin and calbindin, are expressed in largely non-overlapping groups of neurons that show fast-spiking and bursting electrophysiological phenotypes, respectively (Markram et al., 2004).

Parvalbumin is expressed in a group of interneurons characterised by fast responses and effective inhibition of surrounding principal neurons (Hu et al., 2014). The role of parvalbumin neurons in fine-tuning networks of principal neurons has been widely investigated in deep layers of somatosensory and visual cortices (Atallah et al., 2012; Runyan et al., 2010; Yu et al., 2016, 2019). In association cortices, parvalbumin neurons are less prominent in deeper layers, and interestingly, in parahippocampal domains they are primarily seen in superficial layers (Boccara et al., 2015). A well-studied example is the parvalbumin neuron in the medial entorhinal cortex, known for its characteristic grid cells, which have multiple firing fields making up a triangular array across the entire environment available to an animal (Hafting et al., 2005). Parvalbumin interneurons are key modulators of these cells, particularly in layer II networks, where principal neurons communicate through parvalbumin interneurons (Couey et al., 2013; Miao et al., 2017). Similar principles of inhibitory connectivity has been shown in the lateral entorhinal cortex (Nilssen et al., 2018), where principal cells are tuned to the past and present positions of objects (Tsao et al., 2013) and groups of cells are involved in representing sequences of event (Tsao et al., 2018). Beyond the role of parvalbumin neurons in parahippocampal circuits, the importance of these interneurons across the brain is supported by

their dysfunction in several neuropsychiatric and –developmental disorders (Ferguson and Gao, 2018; for review, see Marín, 2012), including autism spectrum disorders (Gogolla et al., 2009), Tourette syndrome (Kalanithi et al., 2005) and schizophrenia (Gonzalez-Burgos and Lewis, 2012; Hashimoto et al., 2003).

Calbindin-D28k is expressed in populations of excitatory and inhibitory neurons (Jinno and Kosaka, 2006; Szabadics et al., 2010). Other calbindin proteins include calbindin-D9k, primarily expressed in epithelial cells, and calretinin, which is expressed in neuronal cells and to a degree co-localizes with calbindin-D28k (Lu et al., 2009; Rogers and Résibois, 1992). In this study, we focus on neurons expressing calbindin-D28k, in the following referred to as calbindin neurons. The functional roles of these neurons have been less well characterised than those of parvalbumin interneurons. However, recent evidence showing that selective knockdown of calbindin neurons in the CA1 and dentate gyrus regions of the hippocampus can reduce long-term potentiation, points to a role for these neurons in memory function (Li et al., 2017). Calbindin neurons have also been implicated in fear memory and social behaviour (Harris et al., 2016), and have been hypothesised to have a neuroprotective role (Sun et al., 2011).

Given the well-known and proposed roles of parvalbumin and calbindin neurons in neuronal networks, quantitative information representing their number and distribution in the brain is of broad interest to neuroscientists. Such data are needed to constrain computational models, to measure group differences in intervention-based studies, and to draw conclusions about structure-function relationships. Several studies have quantified neurons expressing calcium-binding proteins in one or a few brain regions (see, e.g. Andsberg et al., 2001; Pitts et al., 2013; Schmid et al., 2013; Yalcin-Cakmakli et al., 2018). Studies on a larger scale have typically been qualitative or semi-quantitative (Arai et al., 1994; Frantz and Tobin, 1994), while one study has reported brain-wide quantitative data about parvalbumin neurons in Cre reporter mice (Kim et

al., 2017). Others have focused on gathering measurements from the literature and calculating values for parameters that have yet to be tested experimentally (Bezaire and Soltesz, 2013). However, there is growing awareness that numbers reported in the literature are prone to substantial variability across publications (Bjerke et al., 2020a; Keller et al., 2018). Quantitative studies of parvalbumin and calbindin neurons acquired across the brain are needed to elucidate their relative numbers and distributions within and across regions and species. Also, replication and validation of quantitative studies will be essential to converge on realistic estimates of the number of various cell types.

Computational methods for automated segmentation, localisation and quantification of cells have successfully been applied to three-dimensional volumetric datasets to generate region- or brain-wide estimates of cell numbers in mice (Kim et al., 2017; Murakami et al., 2018; Silvestri et al., 2015; Zhang et al., 2017). These efforts have relied on advanced volumetric imaging techniques, genetically modified animals expressing fluorescent signals in cells of interest, and custom codes for analysis. Immunohistochemical techniques continue to serve important purposes for characterising cell populations based on protein expression (which may only be a subset of those expressing the gene for the protein). To achieve efficient quantification of immunohistochemically labelled cells in sectioned material, Yates et al. (2019) combined three open-access tools, QuickNII (Puchades et al., 2019), ilastik (Berg et al., 2019) and Nutil (Groeneboom et al., 2020), in a workflow incorporating customized brain atlas maps, machine learning based segmentation of objects of interest from section images, and quantification of segmented objects in atlas-defined regions of interest.

We here ask how the numbers and spatial distributions of neurons expressing calcium binding proteins vary across brain regions, and possibly relate to functional or topographical patterns of organization. We quantify two largely distinct cell types identified by the calcium binding proteins parvalbumin and calbindin in the mouse brain using the QUINT workflow, and

compare these with previous reports. We further quantify parvalbumin neurons in the entire rat brain, and perform a detailed comparison of parvalbumin neuron numbers in the mouse and rat hippocampal regions. We validate the resulting numbers with manual counts in selected areas, and assess the reliability of segmentation results between researchers. All the raw and derived data sets presented here are shared through the EBRAINS Knowledge Graph to facilitate further analysis and re-use.

RESULTS

We used the QuickNII-ilastik-Nutil (QUINT) workflow to quantify parvalbumin neurons in the mouse and rat brain, and calbindin neurons in the mouse brain, corrected the resulting numbers with Abercrombie's formula, and extrapolated corrected numbers to represent whole regions and volumetric densities. All data were anatomically located using the Allen Mouse Common Coordinate Framework, version 3 of the template, 2017 edition of the delineations (Wang et al., 2020; hereafter referred to as CCFv3-2017) and the Waxholm Space atlas of the Sprague-Dawley rat brain, version 1.01 of the template and version 2 of the delineations (Papp et al., 2014; Kjonigsen et al., 2015; hereafter referred to as WHSv2).

Below, we first present the quantitative data on the densities of parvalbumin and calbindin neurons in the mouse brain (n = 4 and 5, respectively), as density estimates are readily compared across regions of variable sizes. We go on to compare the total numbers of these cell types across the brain and analyse their relative numbers in each brain region. In addition to the text and figures presented here, all numbers and density estimates are listed in Supplemental File 1. We compare our findings to numbers reported in the literature. Additional data describing efforts to assess the validity and reliability of QUINT results are given in Supplemental File 2-4. Lastly, we compare the total number, density and distribution of parvalbumin neurons in the mouse (n = 4) and rat (n = 4) hippocampal regions. All numbers reported are given as mean ± SEM; total number estimates are bilateral, while densities are given per mm^3. The nomenclature used here for mouse anatomical regions follows the CCFv3-2017 hierarchy (except in the cross-species comparison, where WHSv2 terms are used for both species). Some of the overarching terms from the CCFv3-2017 may not be commonly used by researchers; however, these are all listed and explained in Figure 1.

Parvalbumin neuron densities across the mouse brain

Parvalbumin neurons were most densely packed in isocortical areas and in the retrohippocampal region. Olfactory areas, and areas of the hippocampal region, striatum, pallidum and cortical subplate generally showed moderate densities, with some olfactory and amygdalar areas having high densities (for details see below). Low densities were seen in the thalamus and hypothalamus, although some areas stood out with moderate amounts of parvalbumin neurons. Midbrain, pontine and medullary generally showed low and moderate parvalbumin neuron densities. The density estimates for parvalbumin neurons in all grey matter regions of the mouse brain are summarised in Figure 2A, and all total number and density estimates for all mouse brain regions are included in the derived dataset.

Isocortex. Relatively high densities were seen across most isocortical areas. Auditory, visual and somatosensory areas generally showed slightly higher densities than gustatory, visceral, and prefrontal areas. Among isocortical areas, the highest parvalbumin neuron density was seen in the anterolateral visual area (3565 ± 744), while the most sparse distribution was seen in the perirhinal area (999 ± 247).

Olfactory areas. The dorsal peduncular area showed the highest density of the olfactory areas (2240 ± 590), in contrast to a very low density in the accessory olfactory bulb (150 ± 72). Other olfactory areas showed a moderate density.

Hippocampal formation. The hippocampal region showed moderate densities of parvalbumin neurons. In the retrohippocampal regions, densities were generally high, with the parasubiculum showing the highest parvalbumin neuron density of all mouse brain regions (3921 ± 404). A detailed account of parvalbumin neuron densities across the hippocampal formation is given below.

Cortical subplate. Moderate parvalbumin neuron densities were seen across claustrum, entopeduncular and amygdalar regions. The posterior amygdalar nucleus stood out among the cortical subplate regions with a relatively high density (2023 ± 748).

Cerebral nuclei. In the striatum, the dorsal (caudoputamen) part showed a higher density of parvalbumin neurons (866 ± 30) than the ventral (nucleus accumbens and fundus of striatum) region (490 ± 64). Among regions of the pallidum, the medial region showed the highest density (1052 ± 357).

Thalamus and hypothalamus. The thalamic regions generally had low densities of parvalbumin neurons, except for the reticular nucleus (but this region was oversaturated with staining, and numbers are therefore not reported). The geniculate group of the ventral thalamus also showed a relatively high density (809 ± 338). The mammillary body showed the highest density among the hypothalamic regions (1332 ± 258).

Midbrain. In the superior colliculus, a higher density was seen in the superficial (sensory related) part (2505 ± 478) than in deeper (motor related) layers (1095 ± 362). The inferior colliculus (2004 ± 479) and the pretectal region (995 ± 313) also showed relatively high densities compared to other midbrain regions.

Pons and medulla. Pontine and medullary regions were grouped into quite broad categories for the current analysis, but it seemed that the sensory related parts showed higher densities than the motor and behavioural state related ones (pons, sensory related: 1832 ± 115; medulla, sensory related: 2465 ± 240). Parvalbumin staining was seen across cerebellar layers, with Purkinje cell bodies darkly stained, but the staining was oversaturated, preventing extraction of cell numbers and densities.

Calbindin neuron densities across the mouse brain

Calbindin neurons were mapped and quantified throughout the mouse brain. In cortical areas, lightly stained but densely packed cells were typically seen in layer II, while deeper layers had more scattered distribution of strongly stained cells. Density estimates for calbindin neurons in all regions are shown in Figure 2B.

Isocortex. Of the isocortical areas, the infralimbic area showed an especially high density (3506 ± 249), while the frontal pole showed a much more sparse distribution of calbindin neurons (532 ± 219). Within the primary somatosensory area, the trunk region showed the highest density (1582 ± 275) while the mouth region showed the lowest (782 ± 120). Among the auditory and visual areas, densities were quite similar. Among retrosplenial areas, quite distinct differences were seen, with the ventral part having a lower density (306 ± 55) than the dorsal part (803 ± 130). The lateral agranular part of the retrosplenial cortex showed a higher density than both of these (1512 ± 218).

Olfactory areas. Olfactory areas showed a relatively high density of calbindin neurons, with an especially high density seen in the dorsal peduncular area (3324 ± 156). The accessory olfactory bulb, however, showed a very sparse distribution of calbindin neurons (164 ± 48).

Hippocampal formation. The hippocampal areas showed quite low calbindin neuron densities, with the dentate gyrus showed the highest density (876 ± 136). Retrohippocampal regions also generally showed low densities, except for the entorhinal area, lateral part with a moderate density (1659 ± 197) and the hippocampo-amygdalar transition area which showed a relatively high calbindin neuron density (2519 ± 303).

Cortical subplate. The claustrum and entopeduncular nucleus showed relatively high calbindin neuron densities, as did all amygdalar areas, except for the lateral amygdalar nucleus (986 ±

154). The density in the posterior amygdalar nucleus was considerably higher than all other subregions (5745 ± 589) of the cortical subplate.

Cerebral nuclei. Striatal regions generally showed relatively high densities of calbindin neurons, except for the olfactory tubercle which was very sparsely populated (418 ± 48). The caudoputamen had the highest total number of calbindin neurons of all brain regions (48064 ± 7665), but due to the large size of this region, the density was not considerably higher than other regions (1847 ± 295). However, we note that both the caudoputamen and striatum ventral region contained a large population of lightly stained cells, which as mentioned in the methods were not completely represented with our classifier. In the pallidum, quite low density was seen in the globus pallidus, both external (223 ± 24) and internal (57 ± 16) segments, while the ventral and medial regions of the pallidum showed a higher density, especially the latter (2492 ± 471).

Thalamus and hypothalamus. In the thalamus, especially high densities were seen in the subparafascicular nucleus (2474 ± 474) and the subparafascicular area (4964 ± 1526); the nuclei of the medial and midline groups in general had a high density of calbindin neurons. Hypothalamic regions generally showed relatively high densities of calbindin neurons, particularly the ventromedial hypothalamic nucleus (6298 ± 736) and medial preoptic nucleus (8003 ± 975).

Midbrain. The superficial layers of the superior colliculus, grouped under the sensory related superior colliculus in the CCFv3-2017 hierarchy, was densely packed with calbindin neurons (3345 ± 430). This was in contrast to a relatively sparse distribution in the motor related superior colliculus (480 ± 148). In general, midbrain areas were quite lightly stained for calbindin with low cell densities revealed in our analysis, although some areas stood out as densely packed with cells. This included the midbrain trigeminal nucleus (2209 ± 863), ventral tegmental area

(4339 ± 914), paranigral area (2951 ± 620), periaqueductal grey (1270 ± 262), and midbrain raphe nuclei (1872 ± 371).

Pons and medulla. Pontine and medullary regions generally showed low to moderate densities of calbindin neurons. In the cerebellum, intense staining was seen in the Purkinje cells; however, because these cells were much larger than calbindin neurons in the rest of the brain, we did not obtain a satisfactory segmentation and quantitative data are not presented.

Different patterns of parvalbumin and calbindin neuron numbers in the mouse brain

We compared total number estimates of parvalbumin and calbindin neurons across the mouse brain. Results per region for each cell type are shown in Figure 3. This comparison shows that the parvalbumin neurons generally outnumber calbindin neurons in isocortical and retrohippocampal areas. In contrast, the striatal, olfactory and cortical subplate areas generally had higher numbers of calbindin neurons. Striking differences were seen in the thalamus and hypothalamus, where parvalbumin neurons were sparse while calbindin neurons showed high numbers in most of the nuclei. In midbrain areas, and most notably in the inferior colliculus and parabigeminal nucleus, the number of parvalbumin neurons exceeded that of calbindin neurons. In the pons and medulla, parvalbumin neurons showed the highest numbers as well; however, as mentioned, we grouped these regions quite broadly and more fine-grained analysis would be needed to determine if smaller pontine and medullary nuclei might show different ratios of the two cell types.

To further explore the ratios of parvalbumin and calbindin neurons across the brain, we created pie charts showing the total number estimates for each cell type per region in the CCFv3-2017 (Figure 4). The figure clearly shows the trend that parvalbumin neurons were relatively more abundant in isocortical areas, particularly in somatosensory and motor cortical areas. However, in prefrontal cortices, e.g. prelimbic (PL), infralimbic (ILA), and agranular insular (AGA) areas,

247 the balance was more equal (PL) or even shifted towards more calbindin neurons (ILA, AGA).
248 The balance was also more equal in gustatory (GU) and visceral (VISC) cortices, intercalated
249 between somatosensory and olfactory cortical areas. Areas of the hippocampal and
250 retrohippocampal regions also showed higher parvalbumin than calbindin neuron numbers,
251 with the exceptions of the dentate gyrus (DG), lateral entorhinal area (ENTl) and hippocampo-
252 amygdalar transition areas (HATA). In olfactory, striatal and cortical subplate areas, the
253 calbindin neurons were more abundant, with the only exception being the olfactory tubercle
254 (OT) which showed a more balanced ratio. The dorsal pallidal regions had higher parvalbumin
255 numbers, while the ventral and medial pallidum showed higher numbers of calbindin neurons.
256 All nuclei of the thalamus and hypothalamus showed a higher number of calbindin than
257 parvalbumin neurons (except for the geniculate group of the ventral thalamus, where the
258 numbers were similar). In the midbrain, pons and medulla, parvalbumin neurons were generally
259 more abundant than calbindin neurons. However, a balanced ratio was seen in the
260 pedunculopontine nucleus (PPN) and sensory superior colliculus (SCs), while the paranigral
261 area (PN), ventral tegmental area (VTA), midbrain trigeminal nucleus (MEV), periaqueductal
262 grey (PAG) and raphe nuclei (RAmb) stood out with high calbindin neuron numbers relative to
263 parvalbumin neurons.

264 **Comparison of findings to earlier published data**

265 We compared our estimates of parvalbumin neuron densities to the data presented by Kim et al
266 (2017), and to several other published reports (Figure 5). This comparison shows that our
267 parvalbumin density estimates are generally lower than the ones provided by Kim et al (2017).
268 Differences between our estimates and those provided by Kim et al (2017) were on average
269 69% in isocortical areas, 37% in olfactory areas, 18% in hippocampal regions, 36% in the
270 cortical subplate and 30% in the striatum. Larger differences were seen in the globus pallidus,
271 small nuclei of the thalamus and hypothalamus, and in midbrain, pontine and medullary nuclei

where correspondence between atlas nomenclatures was poor. Still, many similar trends in the relative densities across regions can be seen in the datasets. We found 17 studies reporting parvalbumin neuron densities, in regions that could be mapped to closely corresponding region in the CCFv3-2017 atlas. Data for regions that could not be interpreted relative to the CCFv3-2017 atlas were not included in our comparison. 32 estimates of parvalbumin densities were compared with our findings. The majority of these (19 of 32 estimates), mainly from hippocampal and striatal brain regions, were relatively well aligned with our density estimates, while the remaining estimates (13 of 32 estimates were considerably higher than the here reported estimates. All these studies used immunohistochemistry or -fluorescence with antibodies targeting parvalbumin. A more detailed comparison (including information about strains, regions and antibodies used in each study) can be found in Supplemental File 5, together with similar comparison data for rat parvalbumin and mouse calbindin neurons.

Comparative analysis of parvalbumin neurons in the rat and mouse hippocampal region

Parvalbumin neurons play an important role in the spatial circuits of the hippocampal region (Miao et al., 2017), where findings from rats and mice are often used interchangeably. To elucidate whether the parvalbumin neuron population in these species are similarly distributed within and across regions, we acquired immunohistochemical material showing parvalbumin neurons in the rat brain. We here compare the number, densities and distributions of parvalbumin neurons in rat and mouse brain hippocampal regions. We focus on the hippocampal regions as these are relatively similar among the two atlases used here (Kjonigsen et al., 2011; Wang et al., 2020). To facilitate comparison, we grouped regions in the CCFv3-2017 to their corresponding regions in the WHSv2 (Figure 6). The brain regions mentioned in the following section are thus named according to the WHSv2 nomenclature and may differ slightly from CCFv3-2017 region terms used above. As collective terms, we use *hippocampal formation* to refer to regions of the Ammon's horn, dentate gyrus, fasciola cinereum and

subiculum and *parahippocampal region* for the pre- and parasubiculum and the entorhinal, perirhinal and postrhinal cortices. These correspond to the collective terms hippocampal region and retrohippocampal region, respectively, in the CCFv3-2017. We use the term *hippocampal region* to refer to the hippocampal formation and parahippocampal region combined. Note that the medial entorhinal cortex is simply termed «entorhinal cortex» in WHSv2. We refer to it as medial entorhinal cortex in this manuscript, but it is called entorhinal cortex in the shared data files.

Quantitative estimates of parvalbumin neurons in the rat and mouse brain hippocampal regions. Our analysis of the hippocampal regions in the rat (n = 4) showed that the total number of parvalbumin neurons was highest in the medial entorhinal cortex (MEC; 30521 ± 6418), followed by cornu ammonis 1 (CA1; 27 596 ± 1 759), presubiculum (PrS; 23580 ± 1874) and subiculum (Sub; 20706 ± 1 459). The highest density of parvalbumin neurons was seen in the PrS (2535 ± 201) and the parasubiculum (PaS; 1778 ± 188). The lateral entorhinal cortex (LEC) and dentate gyrus (DG) showed the lowest density of all subregions (LEC: 431 ± 93; DG: 347 ± 15). As in the rat, the MEC had the highest number of parvalbumin neurons of all the mouse hippocampal regions (12576 ± 1409), followed by the CA1 (9947 ± 1832) and Sub (8846 ± 816). The density of parvalbumin neurons per mm^3 was highest in the PaS (3921 ± 404) and PrS (3459 ± 519). The least dense parvalbumin neuron population was seen in the LEC (910 ± 112) and DG (267 ± 85). All total number and density estimates for the hippocampal regions of the rat and mouse are summarised in Figure 6.

Thus, the estimated number of parvalbumin neurons was higher across all hippocampal regions of the rat brain as compared to the mouse; an expected finding given the relatively larger brain of the rat. In contrast, the density of parvalbumin neurons was generally higher in the mouse than in the rat (i.e. more parvalbumin neurons per mm^3). Regions within the hippocampal formation were relatively similar in parvalbumin neuron density: compared to mice, rats showed 19 – 38% lower density in regions of Ammon's horn (CA1-3), 3% higher density in

322 the DG and 19% lower in the fasciola cinereum (FC). Larger differences were seen in the
323 parahippocampal regions, particularly in the PaS, the POR and the entorhinal cortices, with rats
324 having 55 – 67% lower density as compared to the mice.

325 The relative density among regions in the hippocampal formation (including in CA1-3, DG,
326 Sub and the FC) was retained across species, with the highest density seen in the FC and Sub,
327 followed by relatively similar densities in CA1-3, and noticeably lower density in the DG. The
328 same was true for the parahippocampal regions, where the PrS and PaS regions showed the
329 highest density of parvalbumin neurons of all hippocampal regions, followed by the MEC,
330 perirhinal areas and POR.

331 *Distribution of parvalbumin within mouse and rat brain hippocampal regions.* Functional
332 gradients are known to exist along the dorsoventral axis of the MEC (Brun et al., 2008;
333 Giocomo et al., 2014), and in hippocampal-parahippocampal connectivity (Strange et al., 2014).
334 We explored whether the density of parvalbumin neurons changed along the dorsoventral axis
335 of each parahippocampal subregion. The results (summarised in Figure 7) indicated that the
336 density of parvalbumin neurons decreases from dorsal to ventral levels in the POR, MEC, LEC
337 and PaS of the rat. We did not observe clear dorsoventral density gradients in the rat PrS or in
338 the PER regions. In the mouse parahippocampal areas, the distribution of parvalbumin neurons
339 decreased in the MEC and LEC and a similar tendency was seen in perirhinal area 35 (PER35).
340 No gradient was seen in the mouse PaS, but a dorsoventral increase in parvalbumin neuron
341 density was observed in the mouse POR. However, this gradient was not seen when excluding
342 the posterolateral visual area. In remaining areas, no clear gradients were observed (Figure 7).

DISCUSSION

We have quantified densities and numbers of immunolabelled calbindin and parvalbumin neurons across the entire mouse brain using the QUINT workflow (Yates et al., 2019), that combines interactive machine learning-based image segmentation with regions-of-interest defined using 3-D anatomical references atlases. We also quantified parvalbumin neurons in the rat brain. We provide the first brain-wide quantitative analysis on the distribution of immunolabelled calbindin and parvalbumin cells. Below, we first discuss the validity and reproducibility of our quantitative results, with particular focus on our comparison to the literature, before elaborating on the quantitative results. Lastly, we discuss how open sharing of the different (raw and derived) components of our datasets may facilitate their re-use in new analyses.

We mapped the brain-wide data on parvalbumin neuron densities provided by Kim et al. (2017) to the regions used in our analysis, and compared the results. The densities obtained by Kim et al. (2017) were well aligned with ours in hippocampal and striatal regions, as well as in many thalamic and hypothalamic regions. Small differences can likely be ascribed to different definitions of the individual areas. Because Kim et al (2017) employed a custom, 3-D reconstructed version of the Allen Reference Atlas (Dong, 2008; Kim et al., 2015), we were not able to reliably compare data from all regions in their atlas version and ours. Significant changes were made to cortical and hippocampal delineations in the CCFv3-2017 delineations (Wang et al., 2020), and nomenclatures differences in brainstem regions indicate that delineations in these regions have changed as well. For isocortical regions, however, Kim and colleagues (2017) reported density estimates that were much higher than ours across almost all regions. Thus, these differences are not likely to be caused only by different definitions of subregions. While we used immunohistochemistry, Kim and colleagues (2017) employed Cre

reporter mice expressing fluorescent protein in parvalbumin neurons. As the two methods will visualise cells that express the parvalbumin protein and parvalbumin gene, respectively, our lower estimates could be caused by cells having only transient production of the protein, e.g. during development (Madisen et al., 2010). Such cells would be detected by the Cre reporter approach used by Kim et al. (2017), but not by immunohistochemistry as used in our study. Differences in segmentation and quantification methods may of course also influence results. Despite differences in data acquisition and analysis, similar trends in the relative densities of cells were seen across brain regions in the two datasets.

We also gathered quantitative estimates from the literature to benchmark our reported values. However, many studies report two- or three-dimensional densities based on one or a few sections that are unlikely to be representative for an entire region. Furthermore, regional areas or volumes estimated from sections will be highly affected by tissue shrinkage occurring during immunohistochemical procedures (Dorph-Petersen et al., 2001), and few report the use of shrinkage correction factors. In our analysis, we used the region volumes from the three-dimensional reference atlas, based on serial two-photon tomography and magnetic resonance imaging templates (for mouse and rat brain atlases, respectively). These are less affected by shrinkage than histological section material, which will result in estimated densities being lower. Given the large effects of tissue shrinkage on density estimates, it has been argued that total number estimates should preferentially be acquired and reported (Oorschot, 1994), but we observe that density estimates are more often reported in the literature. We note that the cell diameter measurement used in Abercrombie's formula in the current study will also be affected by tissue shrinkage, which will to some degree affect our total number estimates. Nevertheless, in the few cases where we found total number estimates from stereological studies with regions of interest closely corresponding to those used in our analysis (Andsberg et al., 2001; Filice et al., 2016; Lauber et al., 2016, 2018), we observed a high degree of correspondence with our

total number estimates. This indicates that our estimates using Abercrombie's formula is not severely biased by the cell diameter approximation used here. Furthermore, for the literature sources that gave estimates from more than one region of interest, the same trend was seen across regions as in our data. We thus believe that the comparison of our parvalbumin data to those provided by Kim et al. (2017) taken together with those found in the other, non-whole brain studies mentioned above indicates that our estimates reveal reproducible trends across regions.

For calbindin neurons in the mouse, very few quantitative estimates were available from the literature, some of which corresponded well to ours and some of which provided much higher numerical values. We note that a subset of calbindin neurons are very lightly stained (in accordance with previous observations, see Frantz & Tobin (1994)). The classifier used in our segmentation successfully extracted calbindin neurons of high and medium staining intensity, but did not extract the most lightly stained neurons, thus our estimates might be considered lower bounds. Lightly stained neurons were seen across the brain but were most abundant in layer II of isocortical areas, in the striatum, the dentate gyrus and hypothalamus. To extract these neurons, it might be necessary to train separate classifiers for different areas, perhaps also using images of higher resolution than used here. However, very light cells may be hard to distinguish from background staining also for a trained neuroanatomist.

We share segmentation results together with the primary data from which they were derived. In addition to increasing transparency of analysis, this facilitates re-use and re-analysis of the derived data. For example, when new versions of the Waxholm Space rat brain atlas or the Allen Mouse Brain atlas are published, our segmentation images can be reanalysed with new atlas maps. Thus, beyond the quantitative derived data presented here, we consider both the primary data and the segmentation maps to be re-usable for the community in the long-term.

Our comparison of calbindin and parvalbumin neuron numbers across the mouse brain revealed largely complementary patterns, possibly indicating differences in the relative contribution of these cell types within regions and across systems. The findings show that parvalbumin neurons are more abundant in motor and sensory areas as well as in most of the hippocampal region, while calbindin neurons are dominant in limbic, thalamic and hypothalamic areas. In the midbrain, parvalbumin neurons were dominant in sensory and motor-related regions such as the inferior colliculus, nucleus sagulum, parabigeminal nucleus, and substantia nigra, while calbindin was more prevalent in regions involved in behavioural state regulation and pain modulation, such as the periaqueductal grey, ventral tegmental area, and midbrain raphe nuclei. These observations provide a neuroanatomical underpinning for recent evidence supporting the importance of calbindin neurons in social and anxiety-like behaviour (Harris et al., 2016) and their susceptibility to stressful events (Li et al., 2017), and supports the already emphasised role of parvalbumin neurons in sensory systems and spatial navigation (Atallah et al., 2012; Miao et al., 2017; Runyan et al., 2010; Yu et al., 2016, 2019).

In our cross-species comparison of parvalbumin neurons in the hippocampal region, we generally found higher parvalbumin neuron densities in mice than in rats, which is consistent with earlier reports of mice having lower total numbers but higher densities of neurons across the brain than rats (Herculano-Houzel et al., 2006). In both species, the density of parvalbumin neurons decreased from dorsal to ventral in the entorhinal cortex. This observation correlates with the increasing scaling of grid cell firing fields along the dorsoventral axis of the MEC (Brun et al., 2008; Stensola et al., 2012). Cells in the LEC have been found to be tuned to object positions and can coordinate to encode time information (Tsao et al., 2013, 2018), although the relationship between their properties and position along the dorsoventral axis is less well defined. A decreasing gradient in the inhibitory input from parvalbumin interneurons has been described from dorsal to ventral in the entorhinal cortex of mice (Beed et al., 2013; Kobro-

441 Flatmoen and Witter, 2019), and dorsoventral gradients in the number of cell bodies have been
442 described qualitatively in the LEC and MEC of adult mice (Fujimaru and Kosaka, 1996). Our
443 evidence of a dorsal to ventral parvalbumin neuron density gradient in the MEC of both rats
444 and mice correlates well with these connectional and functional gradients, and we show similar
445 trends in the LEC as well. Recent research have indeed highlighted the possibility that similar
446 principles govern the microcircuit wiring in MEC and LEC (Nilssen et al., 2018). Our results
447 indicate that a decreasing dorsoventral density of parvalbumin interneurons may be one such
448 principle.

449 Similarly, a decreasing density of parvalbumin neurons was seen from dorsal to ventral in the
450 rat PaS and POR. The mouse POR did not show a dorsoventral decrease in parvalbumin neuron
451 density; if anything, there was an opposite trend with increasing densities from dorsal to ventral.
452 However, caution is warranted when interpreting this result, as no gradient was not seen when
453 excluding the region termed posterolateral visual area in the CCFv3-2017 from our definition
454 of the mouse POR. A dorsoventral decrease in density was seen in the mouse PER35, although
455 this gradient was not as clear as for the other areas mentioned. Thus, several of the
456 parahippocampal regions show a similar trend of decreasing parvalbumin neuron densities from
457 dorsal to ventral, a trend that seemed more wide-spread in the rat.

458 In conclusion, we here present numbers and distributions of parvalbumin and calbindin neurons
459 across the mouse brain. We compare our results to previously published estimates, showing that
460 our estimates of parvalbumin neurons across the mouse brain are well aligned with a previous
461 brain-wide analysis (Kim et al., 2017) and the literature in striatal and hippocampal regions,
462 where several studies have reported quantitative data. However, in other brain regions, larger
463 differences were seen and very few studies were available. Direct comparisons are typically
464 impeded by lack of information in publications, thus highlighting the need for transparent
465 analyses and their reproduction. Furthermore, we compare the number and distribution of

parvalbumin neurons in the mouse hippocampal region with similar data from the rat. Our analysis of parvalbumin and calbindin neurons points to trends within and across brain regions and species that align well with previous studies showing functional and connectional organisation of these cell types.

ACKNOWLEDGEMENTS

We thank Ingrid Reiten and Kasper Kjelsberg for help with the reliability studies; Nicolaas Groeneboom, Grazyna Babinska and Hong Qu for expert technical assistance; and Heidi Kleven and Martin Øvsthus for useful discussions. This work was funded by the European Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreement No. 785907 (Human Brain Project SGA2), Specific Grant Agreement No. 945539 (Human Brain Project SGA3), and The Research Council of Norway under Grant Agreement No. 269774 (INCF Norwegian Node). The experimental work done in Kavli Center was supported by the Research Council of Norway Grant 227769, the Kavli Foundation, the Centre of Excellence scheme of the Research Council of Norway-Centre for Neural Computation Grant 223262, and the National Infrastructure scheme of the Research Council of Norway-NORBRAIN Grant 197467. High-resolution digital images were acquired using infrastructure established through the Norwegian Brain Initiative, NORBRAIN, supported by Research Council of Norway Grant 295721.

AUTHOR CONTRIBUTIONS

IEB conceived the study, performed the immunohistochemical processing and analysis, organized and prepared the related datasets for sharing, and composed the manuscript with comments from all authors. SCY contributed to designing of the data analysis and validity testing and prepared the analytical workflows. AL performed immunohistochemical processing and contributed to organizing related data sets. MPW contributed to supervising the study. MAP supervised the designing of the data analysis and the development of the analytical workflows. JGB supervised the development of the analytical workflows. TBL conceived and supervised the study and the writing of the manuscript. All authors contributed to conceiving the study and the writing the manuscript. All authors reviewed and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests

MAIN FIGURE TITLES AND LEGENDS

Figure 1. Custom regions of interest used for analysis of mouse brain data.

Colour codes and abbreviations for the custom regions used in Nutil Quantifier. These are consistent with the CCFv3-2017 nomenclature, except the three marked with an asterisk. Main titles correspond to high-level regions, while italic subtitles correspond to finer regions.

Figure 2. Parvalbumin and calbindin neuron densities across mouse brain regions.

Bar graph showing the density of parvalbumin (n = 4; A) and calbindin (n = 5, B) neurons across the brain. Error bars indicate SEM. Groups of brain areas are indicated along the x axis. Bars are placed from left to right in the same order as abbreviations are listed and explained in Table 1.

Figure 3. Comparison of total number estimates of parvalbumin and calbindin neurons across the mouse brain.

Bar chart showing bilateral total number estimates of parvalbumin (top panel) and calbindin (bottom panel) neurons in mouse brain regions. Regions are defined and colour coded according to the CCFv3-2017. Error bars indicate SEM. Groups of brain areas are indicated along the x axis. Bars are placed from left to right in the same order as abbreviations are listed and explained in Table 1.

Figure 4. Ratios of parvalbumin and calbindin neurons across the mouse brain.

Pie charts showing the ratio of each cell type (calbindin in light grey, parvalbumin in dark grey) across mouse brain areas. The pie charts are sorted and the background colour coded according to regions in the CCFv3-2017. Abbreviations are detailed in Table 1.

Figure 5. Comparison to previous reports.

A) The parvalbumin neuron density across the mouse brain reported here. B) The parvalbumin neuron density across the mouse brain as reported by Kim et al. (2017). Estimates (n = 32) found in 17 publications (listed below the bar graphs) are plotted in charts A and B according to the most closely matching region in the CCFv3-2017. C) Difference in density estimates reported by Kim et al. (2017) and in the current study (values from current study subtracted from values reported by Kim et al. (2017)). Literature values are plotted according to their difference from the current study (values from the current study subtracted from the values from literature studies). Error bars indicate SEM. Literature references: Andsberg et al., 2001; Ransome and Turnley, 2005; Jinno and Kosaka, 2006; Parrish-Aungst et al., 2007; Smith et al., 2008; Förster, 2008; Moreno-Gonzalez et al., 2009; Neddens and Buonanno, 2009; Pitts et al., 2013; Song et al., 2013; Filice et al., 2016; Lauber et al., 2016, 2018; Fasulo et al., 2017; Grünewald et al., 2017; Pirone et al., 2018; Yalcin-Cakmakli et al., 2018.

Figure 6. Cross-species comparison of parvalbumin neuron numbers and densities in hippocampal regions.

(A) Rostrolateral and caudolateral (with cerebellum and brainstem removed) 3-D views of the hippocampal regions in the WHSv2, shown in the colour they are assigned in the atlas (Kjonigsen et al., 2015) within a transparent view of the brain. (B) Corresponding 3-D views of hippocampal regions in the CCFv3-2017. Regions are colour coded according to their corresponding region in WHSv2. (C) Bar graphs showing the total number of parvalbumin neurons in rat (left) and mouse (right) brain hippocampal regions (D) Bar graph showing the density of parvalbumin neurons in hippocampal regions. Solid bars show rat brain data, patterned bars show mouse brain data. Region names (WHSv2 terms in black text, corresponding CCFv3-2017 term in grey italic text), colour codes according to WHSv2, and

the volume (V) of each region (mouse / rat), are given in the lower panel. Error bars indicate SEM.

Figure 7. Parvalbumin neuron distribution along the dorsoventral axis of hippocampal regions.

(A,B) Coloured bars, each representing one area of the hippocampal region, with individual segments in each bar corresponding to a section along the dorsoventral axis. Approximate Bregma positions of sections along the dorsoventral axis are given. Parvalbumin neuron density is indicated by the intensity of the colour (from light yellow to dark orange). (C) Example images from dorsal and ventral parts of the entorhinal cortex and lateral entorhinal cortex from rat (subject 25205, left panel) and mouse (subject 81266, right panel) showing a denser population of parvalbumin neurons in dorsal parts of these regions. Section numbers and approximate dorsoventral Bregma level is indicated for each image. Abbreviations: POR, postrhinal cortex; PrS, presubiculum; PaS, parasubiculum; MEC, medial entorhinal cortex; PER36, perirhinal area 36; PER35, perirhinal area 35; LEC, lateral entorhinal cortex.

Figure 8. Examples of representative images with atlas overlays and segmentation results.

(A,B) show parvalbumin stained sections (from mouse 81265, sections 012 and 041, respectively), and (C,D) show calbindin stained sections (from mouse 6, sections 097 and 181, respectively). Atlas overlays are superimposed on one hemisphere. (A1-2, B1-2, C1, and D1-3) are enlarged images from selected regions as indicated in (A-D), showing regions secondary motor area (A1, C1), hippocampus (A2, D1), hypothalamus (A3, D2), and piriform area (A4, D3). Panels (A1'-2', B1'-2', C1' and D1'-3') show ilastik segmentations. Some lightly stained calbindin neurons (e.g. in hypothalamus) were not extracted by our classifier. Images are shown at the reduced resolution used for analysis, high-resolution images are available from the shared datasets. Scale bars: upper panels, 1 mm; lower panels, 50 μm.

METHODS

Resource availability

Lead contact

For further information and requests for resources and reagents contact corresponding author, Trygve B. Leergaard (t.b.leergaard@medisin.uio.no).

Materials Availability

This study did not generate new unique reagents.

Data and Code Availability

All raw and derived data from this project are shared via the EBRAINS research infrastructure (https://ebrains.eu/). The primary datasets contain high-resolution TIFF images of the immunohistochemical material, and are shared under the following titles:

1) Distribution of calbindin positive neurons in the normal adult mouse brain (Bjerke and Leergaard, 2020)
2) Distribution of parvalbumin-positive interneurons in the normal adult mouse brain (Laja et al., 2020a)
3) Distribution of parvalbumin-positive interneurons in the normal adult rat brain (Laja et al., 2020b)

The derived datasets contain downscaled PNG images of the primary data, PNG images for the segmentations, atlas maps (PNG and FLAT files), NUT files used to run Nutil Quantifier and all the output reports from this analysis, as well as the final quantitative results per region of interest as presented in this paper. The derived datasets are shared under the following titles and DOIs:

1) Brain-wide quantitative data on calbindin positive neurons in the mouse (Bjerke et al., 2020b)
2) Brain-wide quantitative data on parvalbumin positive neurons in the mouse (Bjerke et al., 2020c)
3) Brain-wide quantitative data on parvalbumin positive neurons in the rat (Bjerke et al., 2020d)

Together, the material provided in the derived dataset allows other researchers to re-run the analysis performed here, re-use the segmentation files with other atlas maps or other parameters in Nutil Quantifier, or re-segment the image material.

Experimental model and subject details

Experimental procedures involving animals were approved by the Norwegian Food Safety Authority and carried out in accordance with the European Union and International legislation for the use of animal subjects. Four adult (6 months old) male PVCre X Rosa26eYFP mice (RRID:IMSR_JAX:008069 and RRID:IMSR_JAX:007903, respectively, crossed locally) and four adult female Sprague-Dawley rats (3 months old, Charles River, Sulzfeld/Kisslegg, Germany) housed at the Kavli Institute, Norwegian Institute of Science and Technology (NTNU), Norway were used for parvalbumin immunohistochemistry. Six adult C57Bl6/J mice (four females and two males, four months old, females from Janvier Labs and males from Taconic), housed at the Institute for Basic Medical Sciences, University of Oslo, Norway, were used for calbindin immunohistochemistry. Animals were housed according to the recommendations by FELASA under pathogen free conditions, with the exception of the presence of Entamoeba muris detected by PCR in fecal samples for the rats. Up to 2 rats and 5 mice of the same sex were group-housed in transparent, semi-enriched cages with 12:12 reversed day / night cycles and ad libitum access to food and water.

METHOD DETAILS

Immunohistochemistry

For parvalbumin immunohistochemistry, animals were deeply anesthetised with sodium pentobarbital (50 mg/kg body weight) and transcardially perfused with Ringer solution containing 4 % paraformaldehyde (PFA). Brains were postfixed in PFA for approximately 24 hours and cryoprotected in a solution of 2 % DMSO and 20 % glycerol in phosphate buffer (PB). Horizontal sections were cut at 30 µm (mouse brains) or 40 µm (rat brains) on a freezing microtome. For calbindin immunohistochemistry, mice were anesthetized with isofluorane, given an overdose of Zoletil mixture, and transcardially perfused with NAPI followed by 4 % PFA. The brains were postfixed overnight in the same fixative, and transferred to 0.4 % PFA for storage. Prior to cutting, the brains were cryoprotected by immersion in 10, 20 and 30 % sucrose until they sank. The right hemisphere was marked by making shallow cut in the cortical surface. Brains were coronally divided at the level of the dorsal hippocampus before coronal sections of 40 µm were cut using a freezing microtome. For all experiments, every sixth section was used for free-floating immunohistochemistry according to the avidin-biotin peroxidase method, using 3,3'-diaminobenzidine (DAB) as the chromogen. All subsequent steps were performed at room temperature, unless otherwise specified.

Parvalbumin immunohistochemistry. Sections were rinsed with PB before blocking endogenous peroxidase activity by incubation in a solution of 3 % H_2O_2 and 10 % methanol in PB. Sections were rinsed again with PB, but subsequent washes were done using Tris buffered saline with 0.5% triton (TBS-TX). The sections were pre-incubated for 1 hour in 5 % normal goat serum in TBS-TX before overnight incubation at 4°C with the primary antibodies (diluted 1:4000 in TBS-TX). A monoclonal mouse anti-parvalbumin antibody (RRID:AB_477329) was used for the rat brain sections, and a monoclonal rabbit anti parvalbumin (RRID:AB_2631173) for the mouse brain sections. After rinsing, sections were incubated overnight at 4°C with

polyclonal, biotinylated secondary antibodies: a goat anti mouse (RRID:AB_258604) antibody for the rat brain sections and a goat anti rabbit (RRID:AB_258649) antibody for the mouse brain sections. Sections were rinsed and then incubated for 1.5 hour with the VectaStain ABC HRP kit, according to manufacturer's instructions. After rinsing again with TBS-TX, the sections were rinsed twice with Tris-HCl before incubation with the DAB solution. The solution was made by adding a 10 mg DAB tablet to 20 mL Tris-HCl, and the tablet was dissolved by placing the solution on a heated stirrer for 2 hours. Afterwards, H_20_2 was added to the solution, which was then placed in the freezer for 20 minutes to slow down the reaction with the sections. The sections were incubated with the DAB solution for approximately 3 minutes and then rinsed with Tris-HCl. The sections were mounted on Superfrost microscope slides with Tris-HCl, and then dried and coverslipped using Entellan (Merck Millipore, Darmstadt, Germany).

Calbindin immunohistochemistry. Sections were rinsed with PBS three times. Endogeneous peroxidase activity was blocked by incubating sections with 3% hydrogen peroxide for five minutes. After three brief rinses in PBS, sections were incubated in blocking solution (3% normal donkey serum, 1% BSA and 0.1% triton X in PBS) for 1 hour. Sections were incubated in the primary antibody (monoclonal mouse anti calbindin, Swant 300, RRID:AB_10000347, lot no. 07F) overnight at 4°C. The antibody is specific to calbindin-D28k and does not bind to calretinin or other known calcium bindin proteins (manufacturer's description). Control sections processed without the primary antibody did not show specific labelling, and are shared together with the data sets. The next day, sections were washed in PBS before incubation with the secondary antibody (sheep anti mouse, GE Healthcare RPN1001V) for one hour. After rinsing, sections were incubated with ABC kit for 30 minutes, and rinsed again with PBS. Sections were then reacted with DAB and H2O2 (kit from Abcam, used according to manufacturer's instructions) for five minutes. The reaction was stopped with distilled water. Sections were mounted from PBS onto gelatinized microscope slides, dehydrated through an

ascending series of ethanol (70, 90 and 100 % for two minutes each) followed by two times two minutes in xylene, dried and coverslipped with Entellan.

Scanning and image pre-processing

Sections were scanned with a 20× objective using a Zeiss Axioscan Z1 scanner (Carl Zeiss MicroImaging, Jena, Germany) and loss less TIFF files exported from the Zen software (RRID:SCR_013672). The TIFF images were renamed to reflect their serial order, and sections that had been horizontally flipped during mounting were mirrored. Images were downscaled to the lowest resolution giving satisfactory segmentation of cell bodies with the ilastik software (15 % and 10 % of their original width for parvalbumin and calbindin images, respectively). Mirroring, renaming and resizing steps were performed using the Transform function of the Nutil software (v.0.4.01, RRID:SCR_017183, Groeneboom et al., 2020). Distortions and dislocation of individual tissue pieces introduced during processing and mounting were corrected using Adobe Photoshop. We moved and mirrored image parts to reconstruct normal section appearance. For parvalbumin sections, such corrections were applied to the original TIFF images. However, the calbindin TIFF images were too large to be opened in Adobe Photoshop or similar software, and correction of section parts were therefore only applied to the downscaled images.

Atlas registration

Serial section images were spatially registered to common 3-D brain reference atlases using the QuickNII software v2.2 (RRID:SCR_016854; Puchades et al., 2019), bundled with the Allen Mouse Common Coordinate Framework, version 3 of the template, 2017 edition of the delineations (Wang et al., 2020; here referred to as CCFv3-2017) and the Waxholm Space atlas of the Sprague-Dawley rat brain, version 1.01 of the template and version 2 of the delineations (Papp et al., 2014; Kjonigsen et al., 2015; here referred to as WHSv2). First, section images

were registered using QuickNII to identify section positions and deviations from the standard planes. For coronally oriented sections, the anatomically distinct landmarks were the genu of the corpus callosum, the crossing of the anterior commissure, and the rostral appearance of structures such as nucleus accumbens, caudoputamen, and dorsal hippocampus. The superior and inferior colliculus, as well as the inferior olive, provided important anatomical landmarks in more posteriorly located sections. For horizontally oriented sections, the registration was primarily based on the anterior commissure, the dorsal appearance of the hippocampus and caudoputamen, the olfactory bulb and the piriform cortex. A second researcher verified all registration results. The affine registration with QuickNII yielded custom atlas plates matching each section image. To further optimize registration, the custom atlas images were non-linearly transformed using the software tool VisuAlign v0.8 (RRID:SCR_017978). We first focused on fitting the template to the outer edges of the section, and secondly adjusted it to refine the fit of major landmarks situated deeper in the brain (e.g. striatum, globus pallidus, hippocampus).

Image segmentation

Pixel classification with ilastik. Images were segmented using the pixel classification pipeline in ilastik, which allows classification of features on a pixel level on a scale up to 10 by 10 pixels. We assigned two label classes termed “cell” and “background”. During segmentation, we placed labels throughout different regions and in all the training sections until the segmentation was deemed satisfactory. We trained one classifier per image series (i.e. one per animal). For each series, we used every fifth image for training and then applied the classifier to all images using the Batch Processing function in ilastik.

Removal of artefacts and incorrectly segmented objects. Overall, the classifiers produced by the pixel classification pipeline distinguished objects of interest from background with relatively high accuracy. Still, some artefacts and incorrectly segmented objects occurred. The

ilastik software includes an object classification pipeline that allows for distinguishing artefacts, which is especially useful when these are very different from the objects of interest. However, in our material incorrectly segmented objects (e.g. neurites) resembled the objects of interest in size and shape. Therefore, we visually inspected all segmentation images to identify incorrectly segmented objects or artefacts, and removed these manually using NIH ImageJ. Most artefacts segmented as "cell" were seen around the edges of the sections and in relation to blood vessels and ventricles. Example images, segmentations and atlas overlays are shown in Figure 8.

Quantification of segmented objects

We used Nutil Quantifier to combine customized atlas maps from QuickNII (refined and exported using VisuAlign) with segmented images from ilastik. The pixel size cut-off function was used to exclude objects representing fragments or small artefacts; a cut-off of 4 pixels was chosen for both parvalbumin and calbindin datasets, based on visual inspection of segmented images.

The regions in the CCFv3-2017 are more fine-grained than the WHSv2. To achieve better correspondence with the granularity of the rat brain atlas, small regions and individual cellular layers in CCFv3-2017 were grouped into more coarse areas (e.g. primary motor cortex, CA1 of the hippocampus) using the "custom regions" feature of Nutil Quantifier. Excel sheets indicating the organization of atlas regions are provided with the derived datasets. Post-processing of numerical data (correction and extrapolation to whole regions and volumetric densities, described below) from mouse brains was performed on the numbers from these aggregated custom regions. The numbers per individual region in the CCFv3-2017 were also extracted with Nutil Quantifier and included in the derived data files. The custom regions used in the present analyses hierarchy, are given in Figure 1.

Post-processing of Nutil Quantifier results

The output reports from Nutil contain the total number of segmented objects counted in every sixth sampled section through each atlas region, given as counts per region, section, and region for the whole brain. These numbers represent all objects counted in all the images investigated. Additionally, damaged or missing parts of individual sections might influence results. To correct for the tissue sampling and missing tissue parts, we implemented a series of post-processing steps to extrapolate Nutil Quantifier numbers to reflect densities and total number estimates. First, we used the sectional reports to compile numbers per section for each region in a new spreadsheet. For all series, we manually inspected the quality and completeness of each section image. Whenever a part of the brain was damaged or missing in a section, or staining quality was deemed suboptimal, we identified which region(s) of the brain this affected. For every such instance, we replaced the Nutil Quantifier results according to the following rules. If a section showed damage or suboptimal staining on one side of the brain, we used the intact side as a reference and multiplied the number by two to get a bilateral estimate for the section. If a region was damaged or sub-optimally stained on both sides, we used an average of numbers obtained from the adjacent neighbouring sections. The adjusted section-by-section numbers were summed per region of interest in a second spreadsheet. On average, 20 % of the sections in a series needed some form of correction, typically meaning that numbers for one or a few regions were corrected as described above. Exclusion of whole sections from the analysis were very rarely necessary, and never for more than 4 % of all the sections within a series. Details on whether corrections were made (and if so, which ones) are provided as metadata with each of the datasets.

Secondly, to adjust for double-counting, we calculated an estimate of total numbers of cells in each region by using Abercrombie's formula (Abercrombie, 1946):

$$N = \frac{n \times T}{T + D}$$

Where N is the cell number, n is the number of counted profiles, T is the thickness of the sections, and D is the mean diameter of the profiles. The mean cell diameter was calculated by multiplying the average object area for all cells (generated by Nutil Quantifier) by the pixel scale of the images, and using this area (A) to calculate the diameter (D):

$$D = 2 \times \frac{A}{\pi}$$

The pixel scales for the section images used in ilastik were 8.6 pixels per μm^2 for rat and mouse parvalbumin section images, and 4.8 pixels per μm^2 for the mouse calbindin section images. The estimated diameter of the cells used in our calculations were 11.3 and 12.7 µm for mouse and rat parvalbumin neurons, respectively, and 7.9 µm for mouse calbindin neurons. Lastly, we multiplied the resulting number with the section sampling fraction (6 in case of all datasets) to estimate the total number of cells per region.

To estimate volumetric densities, we divided the total number of cells for each region with the volumes of each atlas region, exported from the ITK-snap software using the "Volumes and statistics" function. The .xlsx files included in the derived datasets contain all the raw and corrected numbers, as well as the region volumes used for calculations. The volumetric densities estimated here were compared against those reported by Kim et al. (2017) and others in the literature (see section on "Comparison of findings with data from previous reports" below).

Comparative analysis of parvalbumin neurons

To compare the total number and density of parvalbumin neurons in the mouse and rat hippocampal region, hippocampal data from regions in the CCFv3-2017 were grouped to

correspond with rat brain regions defined in the WHSv2 rat brain atlas. Most areas of the hippocampal region were highly comparable between the atlases, in terms of nomenclature, anatomical position and shape of the regions (Figure 6). The regions of Ammon's horn and the dentate gyrus were highly similar. The prosubiculum of the CCFv3-2017, intercalated between the subiculum and the CA1, was grouped as part of the subiculum, while the postsubiculum was included in the presubiculum, as this term is commonly used to refer to the dorsal presubiculum (Taube, 2007; Taube et al., 1990). The perirhinal and ectorhinal cortices in the CCFv3-2017 was considered to correspond to WHSv2 perirhinal area 35 and 36, respectively. Lastly, the postrhinal cortex (POR), was present in both the CCF and WHSv2, but with a much narrower dorsoventral extent in the CCFv3-2017 than the WHSv2. In the WHSv2, the POR lines the caudal pole of the lateral and medial entorhinal cortex dorsally; in the CCFv3-2017, however, the POR does not extend to the most caudal part of the cortex. At this level, a cortical region termed the posterolateral visual area is included in the CCFv3-2017. Anatomically, the combined postrhinal and posterolateral visual area in the CCFv3-2017 resemble the POR in the WHSv2. We therefore included both of these in the postrhinal cortex for our comparative analysis. Figure 6 shows hippocampal regions in the two atlases and the mapping of CCFv3-2017 to WHSv2 regions.

Spatial distribution analysis. To explore and compare possible topographical distribution gradients of parvalbumin neurons in the mouse and rat parahippocampal region, we averaged the section-by-section density of parvalbumin neurons across animals and assessed whether densities of parvalbumin stained neurons changed along the dorsoventral axis of each subregion. To get numbers that could be compared for each level, we calculated volumetric densities per section as follows. First, we calculated corrected numbers of cells for each atlas region per section by applying Abercrombie's formula to the section-wise numbers. Then, for each atlas region, we multiplied the region area in pixels (extracted from Nutil Quantifier

reports) by the pixel scale (see above). For each section, we then divided the corrected number of cells per region by the region area, and divided this by the section thickness (Keller et al., 2018) to get the volumetric density:

$$vN = \frac{n}{T}$$

Where vN is the volumetric density, n is the corrected two-dimensional cell count, and T is the thickness of the section. This gave the density per μm^3; lastly, to get numbers per mm^3, we therefore multiplied these by 10^9.

We next sorted sections according to their dorsoventral level for each animal, and then averaged the density of neurons per level across animals. Each subregion was then represented in a spreadsheet as a row containing the average parvalbumin neuron density for each dorsoventral level (one column per level). We applied conditional formatting to each row so each cell was colour-coded in a gradient according to its value, with the lowest number for each row coded light yellow and the highest number coded dark orange. These colour-coded rows were inspected for cases of clear dorsoventral gradients and copied to Adobe Illustrator to be shown in figures (see Results).

Comparison of findings with previous reports

Kim and colleagues (2017) mapped parvalbumin interneurons across the whole mouse brain using transgenic mice and image segmentation based on convolutional neural networks. They shared all the total number and density estimates through their publication, which we downloaded and used for comparison. They used a custom-built three-dimensional version of the original (two-dimensional) Allen Reference Atlas of the mouse brain (Dong, 2008). The nomenclature used for this version is similar, but not identical, to the one in the version we used

826 in our analysis (CCFv3-2017). We therefore mapped the results found in their files to the
827 custom regions used in our analysis (see above).

828 We also queried the literature to find estimates of the same cell types as those quantified here.
829 To this end, we searched for 1) the RRIDs and catalogue numbers of the antibodies used in our
830 study (using Google Scholar) and 2) articles that mentioned both stereology and "parvalbumin"
831 or "calbindin" in the title or abstract (using PubMed). Details about the search strings used can
832 be found in Supplemental File 6. We supplemented this search with references from a recent
833 review of number estimates in the mouse brain (Keller et al., 2018) and from our previously
834 published database of literature-derived quantitative estimates in the basal ganglia (Bjerke and
835 Leergaard, 2019).

836 *Validity and reliability of segmentations*

837 The results obtained with the QUINT workflow were benchmarked against prior manual counts
838 from both parvalbumin and calbindin stained material that were used as a guide during the
839 segmentation process. We also assessed both intra- and interrater reliability by segmenting
840 material from one of the series (parvalbumin stained sections from rat 25205) several times.
841 Detailed methodology, results and a short discussion of the validity and reliability studies, as
842 well as the data from these analyses, are provided in Supplemental Files 2-4.

843 **QUANTIFICATION AND STATISTICAL ANALYSIS**

844 Quantitative data throughout this paper are reported as mean ± SEM. Summary statistics were
845 extracted using Microsoft Excel (RRID:SCR_016137). Exact values of n, representing the
846 number of animals, are given in the result section.

847

SUPPLEMENTAL ITEM TITLES

Supplemental File 1: Brain-wide total number and density estimates for parvalbumin and calbindin neurons. Related to Figure 2,3,4,5.

Supplemental File 2: Validity and reliability of the QUINT approach. Related to Figure 8.

Supplemental File 3: Documentation for researchers in reliability study. Related to Figure 8.

Supplemental File 4: Data from reliability study. Related to Figure 8.

Supplemental File 5: Details about the literature comparison study. Related to Figure 5.

Supplemental File 6: Literature search strategy. Related to Figure 5.

REFERENCES

Abercrombie, M. (1946). Estimation of nuclear population from microtome sections. Anat. Rec. *94*, 239–247.

Aika, Y., Ren, J.Q., Kosaka, K., and Kosaka, T. (1994). Quantitative analysis of GABA-like-immunoreactive and parvalbumin-containing neurons in the CA1 region of the rat hippocampus using a stereological method, the disector. Exp. Brain Res. *99*, 267–276.

Andsberg, G., Kokaia, Z., and Lindvall, O. (2001). Upregulation of p75 neurotrophin receptor after stroke in mice does not contribute to differential vulnerability of striatal neurons. Exp. Neurol. *169*, 351–363.

Arai, R., Jacobowitz, D., and Deura, S. (1994). Distribution of calretinin, calbindin-D28k, and parvalbumin in the rat thalamus. Brain Res. Bull. *33*, 595–614.

Atallah, B., Bruns, W., Carandini, M., and Scanziani, M. (2012). Parvalbumin-Expressing Interneurons Linearly Transform Cortical Responses to Visual Stimuli. Neuron *73*, 159–170.

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

Baquet, Z., Williams, D., Brody, J., and Smeyne, R. (2009). A comparison of model-based (2D) and design-based (3D) stereological methods for estimating cell number in the substantia nigra pars compacta (SNpc) of the C57BL/6J mouse. Neuroscience *161*, 1082–1090.

Barinka, F., Salaj, M., Rybář, J., Krajčovičová, E., Kubová, H., and Druga, R. (2012). Calretinin, parvalbumin and calbindin immunoreactive interneurons in perirhinal cortex and temporal area Te3V of the rat brain: Qualitative and quantitative analyses. Brain Res. *1436*, 68–80.

878 Beed, P., Gundlfinger, A., Schneiderbauer, S., Song, J., Böhm, C., Burgalossi, A., Brecht, M.,
879 Vida, I., and Schmitz, D. (2013). Inhibitory gradient along the dorsoventral axis in the medial
880 entorhinal cortex. Neuron *79*, 1197–1207.

881 Berg, S., Kutra, D., Kroeger, T., Straehle, C., Kausler, B., Haubold, C., Schiegg, M., Ales, J.,
882 Beier, T., Rudy, M., et al. (2019). Ilastik: Interactive machine learning for (bio)image
883 analysis. Nat. Methods 1226–1232.

884 Berridge, M. (1998). Neuronal calcium signaling. Neuron *21*, 13–26.

885 Bezaire, M.J., and Soltesz, I. (2013). Quantitative assessment of CA1 local circuits:
886 knowledge base for interneuron-pyramidal cell connectivity. Hippocampus *23*, 751–785.

887 Bjerke, I., Øvsthus, M., Andersson, K., Blixhavn, C., Kleven, H., Yates, S., Puchades, M.,
888 Bjaalie, J., and Leergaard, T. (2018). Navigating the murine brain: Toward best practices for
889 determining and documenting neuroanatomical locations in experimental studies. Front.
890 Neuroanat. *12*, 82.

891 Bjerke, I., and Leergaard, T. (2020). Distribution of calbindin positive neurons in the normal
892 adult mouse brain. [Data set] Human Brain Project Neuroinformatics Platform. doi:
893 10.25493/KHNT-KV8

894 Bjerke, I., Puchades, M., Bjaalie, J., and Leergaard, T. (2019). Database of quantitative
895 cellular and subcellular morphological properties from rat and mouse basal ganglia. [Data set]
896 Human Brain Project Neuroinformatics Platform. doi: 10.25493/2FDYXZ-76U

897 Bjerke, I., Puchades, M., Bjaalie, J., and Leergaard, T. (2020a). Database of literature derived
898 cellular measurements from the murine basal ganglia. Sci. Data *7*, 211.

899 Bjerke, I., Yates, S., Puchades, M., Bjaalie, J., and Leergaard, T. (2020b). Brain-wide

quantitative data on calbindin positive neurons in the mouse. [Data set] EBRAINS. doi:10.25493/TT2Y-23N

Bjerke, I., Yates, S., Puchades, M., Bjaalie, J., and Leergaard, T. (2020c). Brain-wide quantitative data on parvalbumin positive neurons in the mouse. [Data set] EBRAINS. doi:10.25493/BT8X-FN9

Bjerke, I., Yates, S., Puchades, M., Bjaalie, J., and Leergaard, T. (2020d). Brain-wide quantitative data on parvalbumin positive neurons in the rat. [Data set] EBRAINS. doi:10.25493/KR92-C33

Boccara, C.N., Kjonigsen, L.J., Hammer, I.M., Bjaalie, J.G., Leergaard, T.B., and Witter, M.P. (2015). A three-plane architectonic atlas of the rat hippocampal region. Hippocampus *25*, 838–857.

Brun, V.H., Solstad, T., Kjelstrup, K.B., Fyhn, M., Witter, M.P., Moser, E.I., and Moser, M.B. (2008). Progressive increase in grid scale from dorsal to ventral medial entorhinal cortex. Hippocampus *18*, 1200–1212.

Couey, J.J., Witoelar, A., Zhang, S.J., Zheng, K., Ye, J., Dunn, B., Czajkowski, R., Moser, M.B., Moser, E.I., Roudi, Y., et al. (2013). Recurrent inhibitory circuitry as a mechanism for grid formation. Nat. Neurosci. *16*, 318–324.

Dong, H. (2008). Allen Reference Atlas: A digital color brain atlas of the C57BL/6J male mouse (Hoboken, NJ: John Wiley & Sons).

Dorph-Petersen, K.-A., Nyengaard, J.R., and Gundersen, H.J.G. (2001). Tissue shrinkage and unbiased stereological estimation of particle number and size. J. Microsc. *204*, 232–246.

Fasulo, L., Brandi, R., Arisi, I., La Regina, F., Berretta, N., Capsoni, S., D'Onofrio, M., and

Cattaneo, A. (2017). ProNGF drives localized and cell selective parvalbumin interneuron and perineuronal net depletion in the dentate gyrus of transgenic mice. Front. Mol. Neurosci. *10*, 20.

Ferguson, B., and Gao, W. (2018). PV interneurons: Critical regulators of E/I balance for prefrontal cortex-dependent behavior and psychiatric disorders. Front. Neural Circuits *12*.

Filice, F., Vörckel, K.J., Sungur, A.Ö., Wöhr, M., and Schwaller, B. (2016). Reduction in parvalbumin expression not loss of the parvalbumin-expressing GABA interneuron subpopulation in genetic parvalbumin and shank mouse models of autism. Mol. Brain 1–17.

Förster, J. (2008). Quantitative morphological analysis of the neostriatum and the cerebellum of tenascin-C deficient mice. University of Hamburg.

Frantz, G., and Tobin, A. (1994). Cellular distribution of calbindin D28K mRNAs in the adult mouse brain. J. Neurosci. Res. *37*, 287–302.

Fujimaru, Y., and Kosaka, T. (1996). The distribution of two calcium binding proteins, calbindin D-28K and parvalbumin, in the entorhinal cortex of the adult mouse. Neurosci. Res.

Giocomo, L., Stensola, T., Bonnevie, T., Van Cauter, T., Moser, M., and Moser, E. (2014). Topography of head direction cells in medial entorhinal cortex. Curr. Biol. *24*, 252–262.

Gogolla, N., LeBlanc, J., Quast, K., Südhof, T., Fagiolini, M., and Hensch, T. (2009). Common circuit defect of excitatory-inhibitory balance in mouse models of autism. J. Neurodev. Disord. *1*, 172–181.

Gonzalez-Burgos, G., and Lewis, D. (2012). NMDA receptor hypofunction, parvalbumin-positive neurons, and cortical gamma oscillations in schizophrenia. Schizophr. Bull. *38*, 950–957.

944 Grünewald, B., Lange, M.D., Werner, C., O'Leary, A., Weishaupt, A., Popp, S., Pearce, D.A.,
945 Wiendl, H., Reif, A., Pape, H.C., et al. (2017). Defective synaptic transmission causes disease
946 signs in a mouse model of juvenile neuronal ceroid lipofuscinosis. Elife *6*.

947 Hafting, T., Fyhn, M., Molden, S., Moser, M.B., and Moser, E.I. (2005). Microstructure of a
948 spatial map in the entorhinal cortex. Nature *436*, 801–806.

949 Harris, E., Abel, J., Tejada, L., and Rissman, E. (2016). Calbindin knockout alters sex-specific
950 regulation of behavior and gene expression in amygdala and prefrontal cortex. Endocrinology
951 *157*, 1967–1979.

952 Hashimoto, T., Volk, D., Eggan, S., Mirnics, K., Pierri, J., Sun, Z., Sampson, A., and Lewis,
953 D. (2003). Gene expression deficits in a subclass of GABA neurons in the prefrontal cortex of
954 subjects with schizophrenia. J. Neurosci. *23*, 6315–6326.

955 Hedreen, J.C. (1998). Lost caps in histological counting methods. Anat. Rec. *250*, 366–372.

956 Herculano-Houzel, S., Mota, B., and Lent, R. (2006). Cellular scaling rules for rodent brains.
957 Proc. Natl. Acad. Sci. U. S. A. *103*, 12138–12143.

958 Hu, H., Gan, J., and Jonas, P. (2014). Fast-spiking, parvalbumin^{+} GABAergic interneurons:
959 From cellular design to microcircuit function. Science (80-.). *345*.

960 Jinno, S., and Kosaka, T. (2006). Cellular architecture of the mouse hippocampus: A
961 quantitative aspect of chemically defined GABAergic neurons with stereology. Neurosci. Res.
962 *56*, 229–245.

963 Kaalund, S.S., Riise, J., Broberg, B. V., Fabricius, K., Karlsen, A.S., Secher, T., Plath, N., and
964 Pakkenberg, B. (2013). Differential expression of parvalbumin in neonatal phencyclidine-treated rats
965 and socially isolated rats. J. Neurochem. *124*, 548–557.

966 Kalanithi, P., Zheng, W., Kataoka, Y., DiFiglia, M., Grantz, H., Saper, C., Schwartz, M.,
967 Leckman, J., and Vaccarino, F. (2005). Altered parvalbumin-positive neuron distribution in
968 basal ganglia of individuals with Tourette syndrome. Proc. Natl. Acad. Sci. U. S. A. *102*,
969 13307–13312.

970 Keller, D., Erö, C., and Markram, H. (2018). Cell densities in the mouse brain: A systematic
971 review. Front. Neuroanat. *12*, 83.

972 Kim, Y., Venkataraju, K.U., Pradhan, K., Mende, C., Taranda, J., Turaga, S.C., Arganda-
973 Carreras, I., Ng, L., Hawrylycz, M.J., Rockland, K.S., et al. (2015). Mapping social behavior-
974 induced brain activation at cellular resolution in the mouse. Cell Rep. *10*, 292–305.

975 Kim, Y., Yang, G., Pradhan, K., Venkataraju, K., Bota, M., García del Molino, L., Fitzgerald,
976 G., Ram, K., He, M., Levine, J., et al. (2017). Brain-wide maps reveal stereotyped cell-type-
977 based cortical architecture and subcortical sexual dimorphism. Cell *171*, 456–469.

978 Kjonigsen, L., Lillehaug, S., Bjaalie, J., Witter, M., and Leergaard, T. (2015). Waxholm
979 Space atlas of the rat brain hippocampal region: Three-dimensional delineations based on
980 magnetic resonance and diffusion tensor imaging. Neuroimage *108*, 441–449.

981 Kjonigsen, L.J., Leergaard, T.B., Witter, M.P., and Bjaalie, J.G. (2011). Digital atlas of
982 anatomical subdivisions and boundaries of the rat hippocampal region. Front. Neuroinform. *5*,
983 2.

984 Kobro- Flatmoen, A., and Witter, M. (2019). Neuronal chemo- architecture of the entorhinal
985 cortex: A comparative review. Eur. J. Neurosci. *50*, 3627–3662.

986 Laja, A., Bjerke, I., Leergaard, T., and Witter, M. (2020a). Distribution of parvalbumin-
987 positive interneurons in the normal adult mouse brain. [Data set] EBRAINS.

988 doi:10.25493/BXGX-WM4

989 Laja, A., Bjerke, I., Leergaard, T., and Witter, M. (2020b). Distribution of parvalbumin-
990 positive interneurons in the normal adult rat brain. [Data set] EBRAINS.
991 doi:10.25493/8KCQ-3C7

992 Lauber, E., Filice, F., and Schwaller, B. (2016). Prenatal Valproate Exposure Differentially
993 Affects Parvalbumin-Expressing Neurons and Related Circuits in the Cortex and Striatum of
994 Mice. Front. Mol. Neurosci. *9*, 1–16.

995 Lauber, E., Filice, F., and Schwaller, B. (2018). Dysregulation of Parvalbumin Expression in
996 the Cntnap2 − / − Mouse Model of Autism Spectrum Disorder. Front. Mol. Neurosci. *11*, 1–
997 15.

998 Li, J., Xie, X., Yu, J., Sun, Y., Liao, X., Wang, X., Su, Y., Liu, Y., Schmidt, M., Wang, X., et
999 al. (2017). Suppressed Calbindin Levels in Hippocampal Excitatory Neurons Mediate Stress-
1000 Induced Memory Loss. Cell Rep.

1001 Lu, E., Llano, D.A., and Sherman, S.M. (2009). Different distributions of calbindin and
1002 calretinin immunostaining across the medial and dorsal divisions of the mouse medial
1003 geniculate body. Hear. Res. *257*, 16–23.

1004 Madisen, L., Zwingman, T.A., Sunkin, S.M., Oh, S.W., Zariwala, H.A., Gu, H., Ng, L.L.,
1005 Palmiter, R.D., Hawrylycz, M.J., Jones, A.R., et al. (2010). A robust and high-throughput Cre
1006 reporting and characterization system for the whole mouse brain. Nat. Neurosci. *13*, 133–140.

1007 Marín, O. (2012). Interneuron dysfunction in psychiatric disorders. Nat. Rev. Neurosci. *13*,
1008 107–120.

1009 Markram, H., Toledo-Rodriguez, M., Wang, Y., Gupta, A., Silberberg, G., and Wu, C. (2004).

1010 Interneurons of the neocortical inhibitory system. Nat. Rev. Neurosci. *5*, 793–807.

1011 Megahed, T., Hattiangady, B., Shuai, B., and Shetty, A.K. (2015). Parvalbumin and neuropeptide Y
1012 expressing hippocampal GABA-ergic inhibitory interneuron numbers decline in a model of Gulf War
1013 illness. Front. Cell. Neurosci. *8*, 1–12.

1014 Miao, C., Cao, Q., Moser, M., and Moser, E. (2017). Parvalbumin and Somatostatin
1015 Interneurons Control Different Space-Coding Networks in the Medial Entorhinal Cortex. Cell
1016 *171*, 507-521.e17.

1017 Moreno-Gonzalez, I., Baglietto-Vargas, D., Sanchez-Varo, R., Jimenez, S., Trujillo-Estrada,
1018 L., Sanchez-Mejias, E., Del Rio, J.C., Torres, M., Romero-Acebal, M., Ruano, D., et al.
1019 (2009). Extracellular amyloid-β and cytotoxic glial activation induce significant entorhinal
1020 neuron loss in young PS1M146L/APP751SL mice. J. Alzheimer's Dis. *18*, 755–776.

1021 Murakami, T., Mano, T., Saikawa, S., Horiguchi, S., Shigeta, D., Baba, K., Sekiya, H.,
1022 Shimizu, Y., Tanaka, K., Kiyonari, H., et al. (2018). A three-dimensional single-cell-
1023 resolution whole-brain atlas using CUBIC-X expansion microscopy and tissue clearing. Nat.
1024 Neurosci. *21*, 625.

1025 Neddens, J., and Buonanno, A. (2009). Selective populations of hippocampal interneurons
1026 express ErbB4 and their number and distribution is altered in ErbB4 knockout mice.
1027 Hippocampus *20*.

1028 Nilssen, E., Jacobsen, B., Fjeld, G., Nair, R., Blankvoort, S., Kentros, C., and Witter, M.
1029 (2018). Inhibitory connectivity dominates the fan cell network in layer II of lateral entorhinal
1030 cortex. J. Neurosci. *38*, 9712–9727.

1031 Oorschot, D. (1994). Are you using neuronal densities, synaptic densities or neurochemical
1032 densities as your definitive data? There is a better way to go. Prog. Neurobiol. *44*, 233–247.

1033 Papp, E., Leergaard, T., Calabrese, E., Johnson, G., and Bjaalie, J. (2014). Waxholm Space
1034 atlas of the Sprague Dawley rat brain. Neuroimage *97*, 374–386.

1035 Parrish-Aungst, S., Shipley, M.T., Erdelyi, F., Szabo, G., and Puche, A.C. (2007).
1036 Quantitative analysis of neuronal diversity in the mouse olfactory bulb. J. Comp. Neurol. *501*,
1037 825–836.

1038 Pirone, A., Alexander, J.M., Koenig, J.B., Cook-Snyder, D.R., Palnati, M., Wickham, R.J.,
1039 Eden, L., Shrestha, N., Reijmers, L., Biederer, T., et al. (2018). Social Stimulus Causes
1040 Aberrant Activation of the Medial Prefrontal Cortex in a Mouse Model With Autism-Like
1041 Behaviors. Front. Synaptic Neurosci. *10*, 35.

1042 Pitts, M.W., Reeves, M.A., Hashimoto, A.C., Ogawa, A., Kremer, P., Seale, L.A., and Berry,
1043 M.J. (2013). Deletion of selenoprotein M leads to obesity without cognitive deficits. J. Biol.
1044 Chem. *288*, 26121–26134.

1045 Puchades, M., Csucs, G., Ledergerber, D., Leergaard, T., and Bjaalie, J. (2019). Spatial
1046 registration of serial microscopic brain images to three-dimensional reference atlases with the
1047 QuickNII tool. PLoS One *14*.

1048 Ransome, M.I., and Turnley, A.M. (2005). Analysis of neuronal subpopulations in mice over-
1049 expressing suppressor of cytokine signaling-2. Neuroscience *132*, 673–687.

1050 Rogers, J.H., and Résibois, A. (1992). Calretinin and calbindin-D28k in rat brain: Patterns of
1051 partial co-localization. Neuroscience *51*, 843–865.

1052 Runyan, C.A., Schummers, J., Van Wart, A., Kuhlman, S.J., Wilson, N.R., Huang, Z.J., and
1053 Sur, M. (2010). Response Features of Parvalbumin-Expressing Interneurons Suggest Precise
1054 Roles for Subtypes of Inhibition in Visual Cortex. Neuron *67*, 847–857.

Schmid, J.S., Bernreuther, C., Nikonenko, A.G., Ling, Z., Mies, G., Hossmann, K.A., Jakovcevski, I., and Schachner, M. (2013). Heterozygosity for the mutated X-chromosome-linked L1 cell adhesion molecule gene leads to increased numbers of neurons and enhanced metabolism in the forebrain of female carrier mice. Brain Struct. Funct. *218*, 1375–1390.

Schwaller, B. (2010). Cytosolic Ca2+ buffers. Cold Spring Harb. Perspect. Biol. *2*, a004051.

Shiraki, A., Tanaka, T., Watanabe, Y., Saito, F., Akahori, Y., Imatanaka, N., Yoshida, T., and Shibutani, M. (2016). Immunohistochemistry of aberrant neuronal development induced by 6-propyl-2-thiouracil in rats. Toxicol. Lett. *261*, 59–71.

Silvestri, L., Paciscopi, M., Soda, P., Biamonte, F., Iannello, G., Frasconi, P., and Pavone, F. (2015). Quantitative neuroanatomy of all Purkinje cells with light sheet microscopy and high-throughput image analysis. Front. Neuroanat. *9*.

Smith, K.M., Fagel, D.M., Stevens, H.E., Maragnoli, M.E., Rabenstein, R.L., Ohkubo, Y., Picciotto, M.R., Schwartz, M.L., and Vaccarino, F.M. (2008). Deficiency in Inhibitory Cortical Interneurons Associates with Hyperactivity in Fibroblast Growth Factor Receptor 1 Mutant Mice. Biol. Psychiatry *63*, 953–962.

Song, C.-H., Bernhard, D., Bolarinwa, C., Hess, E.J., Smith, Y., and Jinnah, H.A. (2013). Subtle microstructural changes of the striatum in a DYT1 knock-in mouse model of dystonia. Neurobiol. Dis. *54*, 362–371.

Stensola, H., Stensola, T., Solstad, T., Frøland, K., Moser, M., and Moser, E. (2012). The entorhinal grid map is discretized. Nature *492*, 72–78.

Strange, B.A., Witter, M.P., Lein, E.S., and Moser, E.I. (2014). Functional organization of the hippocampal longitudinal axis. Nat. Rev. Neurosci. *15*, 655–669.

1077 Sun, S., Li, F., Gao, X., Zhu, Y., Chen, J., Zhu, X., Yuan, H., and Gao, D. (2011). Calbindin-
1078 D28K inhibits apoptosis in dopaminergic neurons by activation of the PI3-kinase-Akt
1079 signaling pathway. Neuroscience.

1080 Szabadics, J., Varga, C., Brunner, J., Chen, K., and Soltesz, I. (2010). Granule cells in the
1081 CA3 area. J. Neurosci. *30*, 8296–8307.

1082 Taube, J. (2007). The head direction signal: Origins and sensory-motor integration. Annu.
1083 Rev. Neurosci. *30*, 181–207.

1084 Taube, J.S., Muller, R.U., and Ranck, J.B. (1990). Head-direction cells recorded from the
1085 postsubiculum in freely moving rats. II. Effects of environmental manipulations. J. Neurosci.
1086 *10*, 436–447.

1087 Tsao, A., Moser, M.-B., and Moser, E.I. (2013). Traces of Experience in the Lateral
1088 Entorhinal Cortex. Curr. Biol. *23*, 399–405.

1089 Tsao, A., Sugar, J., Lu, L., Wang, C., Knierim, J., Moser, M., and Moser, E. (2018).
1090 Integrating time from experience in the lateral entorhinal cortex. Nature *561*, 57–62.

1091 Wang, C.Z., Yang, S.F., Xia, Y., and Johnson, K.M. (2008). Postnatal phencyclidine administration
1092 selectively reduces adult cortical parvalbumin-containing interneurons. Neuropsychopharmacology *33*,
1093 2442–2455.

1094 Wang, Q., Ding, S., Li, Y., Royall, J., Feng, D., Lesnar, P., Graddis, N., Naeemi, M., Facer,
1095 B., Ho, A., et al. (2020). The Allen Mouse Brain Common Coordinate Framework: A 3D
1096 reference atlas. Cell *181*, 1–18.

1097 Yalcin-Cakmakli, G., Rose, S.J., Villalba, R.M., Williams, L., Jinnah, H.A., Hess, E.J., and
1098 Smith, Y. (2018). Striatal Cholinergic Interneurons in a Knock-in Mouse Model of L -DOPA-

Responsive Dystonia. Front. Syst. Neurosci. *12*, 1–12.

Yates, S., Groeneboom, N., Coello, C., Lichtenthaler, S., Kuhn, P.-H., Demuth, H.-U., Hartlage-Rübsamen, M., Roßner, S., Leergaard, T., Kreshuk, A., et al. (2019). QUINT: Workflow for Quantification and Spatial Analysis of Features in Histological Images From Rodent Brain. Front. Neuroinform. *13*, 75.

Yu, J., Gutnisky, D.A., Hires, S.A., and Svoboda, K. (2016). Layer 4 fast-spiking interneurons filter thalamocortical signals during active somatosensation. Nat. Neurosci. *19*, 1647–1657.

Yu, J., Hu, H., Agmon, A., and Svoboda, K. (2019). Recruitment of GABAergic Interneurons in the Barrel Cortex during Active Tactile Behavior. Neuron *104*, 412-427.e4.

Zhang, C., Yan, C., Ren, M., Li, A., Quan, T., and Gong, H. (2017). A platform for stereological quantitative analysis of the brain- wide distribution of type-specific neurons. Sci. Rep. *7*, 1–12.

Figure 1

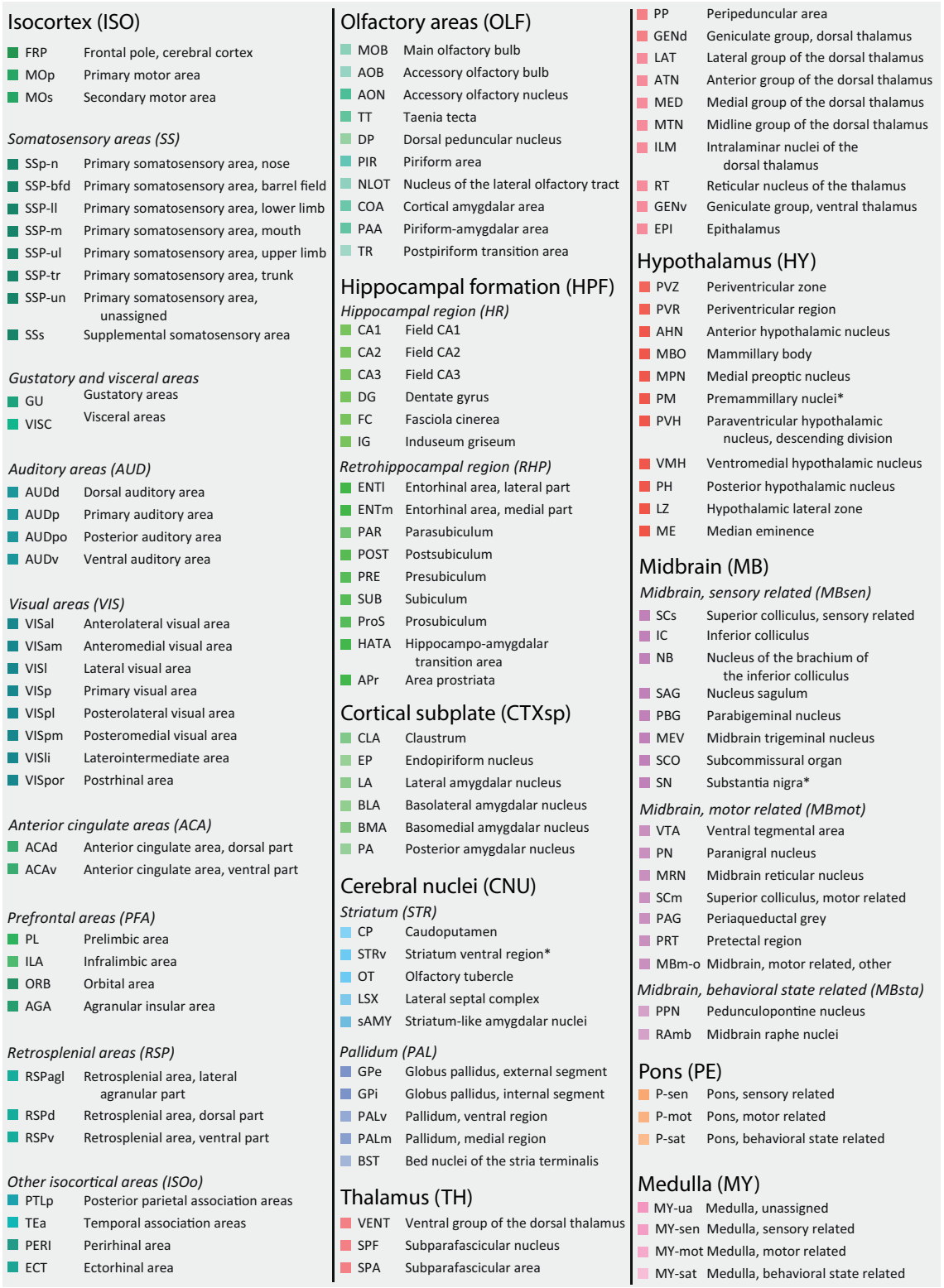

Isocortex (ISO)
FRP Frontal pole, cerebral cortex
MOp Primary motor area
MOs Secondary motor area
Somatosensory areas (SS)
SSp-n Primary somatosensory area, nose
SSP-bfd Primary somatosensory area, barrel field
SSP-ll Primary somatosensory area, lower limb
SSP-m Primary somatosensory area, mouth
SSP-ul Primary somatosensory area, upper limb
SSP-tr Primary somatosensory area, trunk
SSP-un Primary somatosensory area, unassigned
SSs Supplemental somatosensory area
Gustatory and visceral areas
GU Gustatory areas
VISC Visceral areas
Auditory areas (AUD)
AUDd Dorsal auditory area
AUDp Primary auditory area
AUDpo Posterior auditory area
AUDv Ventral auditory area
Visual areas (VIS)
VISal Anterolateral visual area
VISam Anteromedial visual area
VISl Lateral visual area
VISp Primary visual area
VISpl Posterolateral visual area
VISpm Posteromedial visual area
VISli Laterointermediate area
VISpor Postrhinal area
Anterior cingulate areas (ACA)
ACAd Anterior cingulate area, dorsal part
ACAv Anterior cingulate area, ventral part
Prefrontal areas (PFA)
PL Prelimbic area
ILA Infralimbic area
ORB Orbital area
AGA Agranular insular area
Retrosplenial areas (RSP)
RSPagl Retrosplenial area, lateral agranular part
RSPd Retrosplenial area, dorsal part
RSPv Retrosplenial area, ventral part
Other isocortical areas (ISOo)
PTLp Posterior parietal association areas
TEa Temporal association areas
PERI Perirhinal area
ECT Ectorhinal area
Olfactory areas (OLF)
MOB Main olfactory bulb
AOB Accessory olfactory bulb
AON Accessory olfactory nucleus
TT Taenia tecta
DP Dorsal peduncular nucleus
PIR Piriform area
NLOT Nucleus of the lateral olfactory tract
COA Cortical amygdalar area
PAA Piriform-amygdalar area
TR Postpiriform transition area
Hippocampal formation (HPF)
Hippocampal region (HR)
CA1 Field CA1
CA2 Field CA2
CA3 Field CA3
DG Dentate gyrus
FC Fasciola cinerea
IG Induseum griseum
Retrohippocampal region (RHP)
ENTl Entorhinal area, lateral part
ENTm Entorhinal area, medial part
PAR Parasubiculum
POST Postsubiculum
PRE Presubiculum
SUB Subiculum
ProS Prosubiculum
HATA Hippocampo-amygdalar transition area
APr Area prostriata
Cortical subplate (CTXsp)
CLA Claustrum
EP Endopiriform nucleus
LA Lateral amygdalar nucleus
BLA Basolateral amygdalar nucleus
BMA Basomedial amygdalar nucleus
PA Posterior amygdalar nucleus
Cerebral nuclei (CNU)
Striatum (STR)
CP Caudoputamen
STRv Striatum ventral region*
OT Olfactory tubercle
LSX Lateral septal complex
sAMY Striatum-like amygdalar nuclei
Pallidum (PAL)
GPe Globus pallidus, external segment
GPi Globus pallidus, internal segment
PALv Pallidum, ventral region
PALm Pallidum, medial region
BST Bed nuclei of the stria terminalis
Thalamus (TH)
VENT Ventral group of the dorsal thalamus
SPF Subparafascicular nucleus
SPA Subparafascicular area
PP Peripeduncular area
GENd Geniculate group, dorsal thalamus
LAT Lateral group of the dorsal thalamus
ATN Anterior group of the dorsal thalamus
MED Medial group of the dorsal thalamus
MTN Midline group of the dorsal thalamus
ILM Intralaminar nuclei of the dorsal thalamus
RT Reticular nucleus of the thalamus
GENv Geniculate group, ventral thalamus
EPI Epithalamus
Hypothalamus (HY)
PVZ Periventricular zone
PVR Periventricular region
AHN Anterior hypothalamic nucleus
MBO Mammillary body
MPN Medial preoptic nucleus
PM Premammillary nuclei*
PVH Paraventricular hypothalamic nucleus, descending division
VMH Ventromedial hypothalamic nucleus
PH Posterior hypothalamic nucleus
LZ Hypothalamic lateral zone
ME Median eminence
Midbrain (MB)
Midbrain, sensory related (MBsen)
SCs Superior colliculus, sensory related
IC Inferior colliculus
NB Nucleus of the brachium of the inferior colliculus
SAG Nucleus sagulum
PBG Parabigeminal nucleus
MEV Midbrain trigeminal nucleus
SCO Subcommissural organ
SN Substantia nigra*
Midbrain, motor related (MBmot)
VTA Ventral tegmental area
PN Paranigral nucleus
MRN Midbrain reticular nucleus
SCm Superior colliculus, motor related
PAG Periaqueductal grey
PRT Pretectal region
MBm-o Midbrain, motor related, other
Midbrain, behavioral state related (MBsta)
PPN Pedunculopontine nucleus
RAmb Midbrain raphe nuclei
Pons (PE)
P-sen Pons, sensory related
P-mot Pons, motor related
P-sat Pons, behavioral state related
Medulla (MY)
MY-ua Medulla, unassigned
MY-sen Medulla, sensory related
MY-mot Medulla, motor related
MY-sat Medulla, behavioral state related

Figure 2

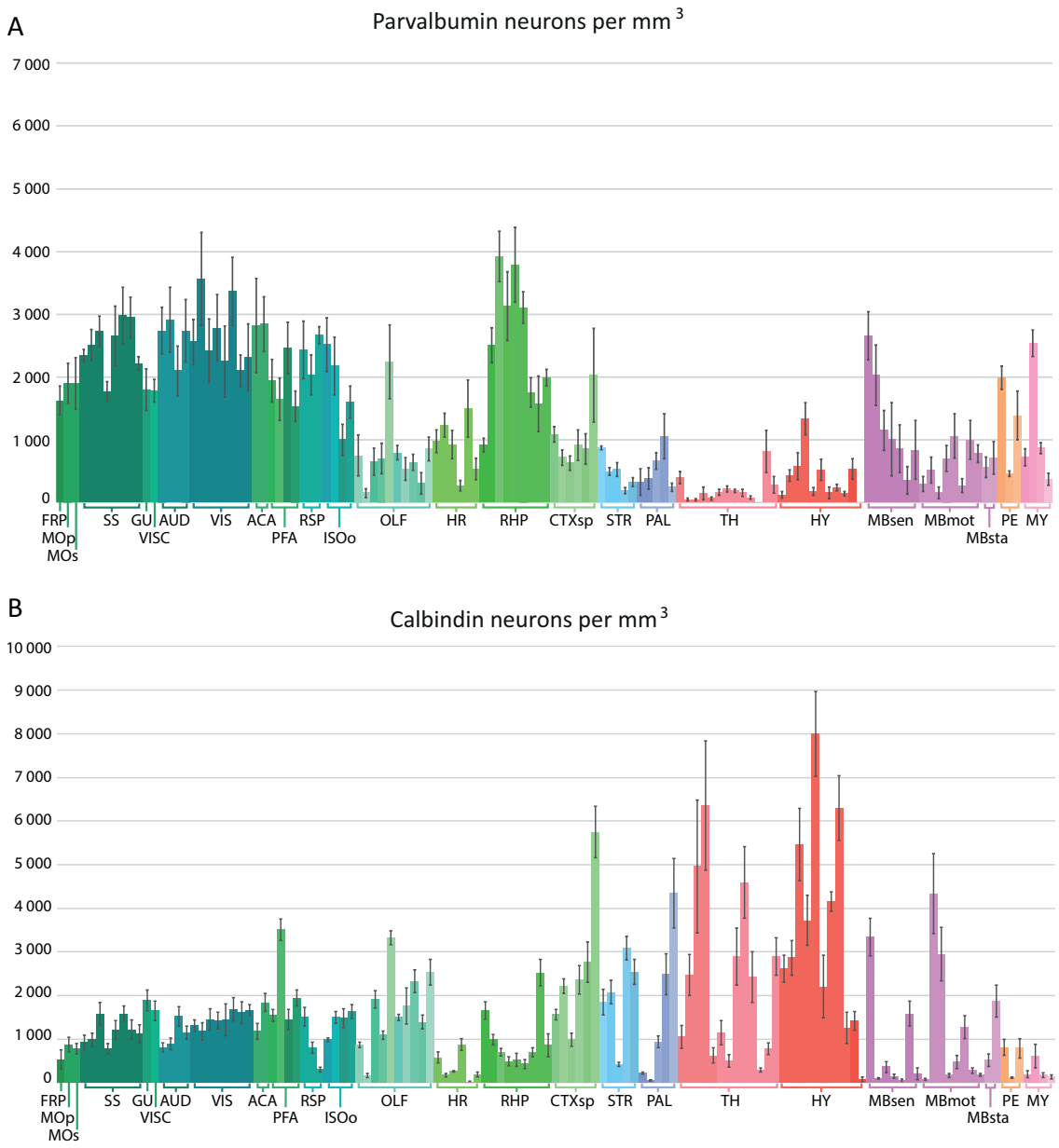

A
Parvalbumin neurons per mm 3
7 000
6 000
5 000
4 000
3 000
2 000
1 000
0
FRP
MOp
MOs
SS
GU
VISC
AUD
VIS
ACA
PFA
RSP
ISOo
OLF
HR
RHP
CTXsp
STR
PAL
TH
HY
MBsen
MBmot
MBsta
PE
MY
B
Calbindin neurons per mm 3
10 000
9 000
8 000
7 000
6 000
5 000
4 000
3 000
2 000
1 000
0
FRP
MOp
MOs
SS
GU
VISC
AUD
VIS
ACA
PFA
RSP
ISOo
OLF
HR
RHP
CTXsp
STR
PAL
TH
HY
MBsen
MBmot
MBsta
PE
MY

Figure 3

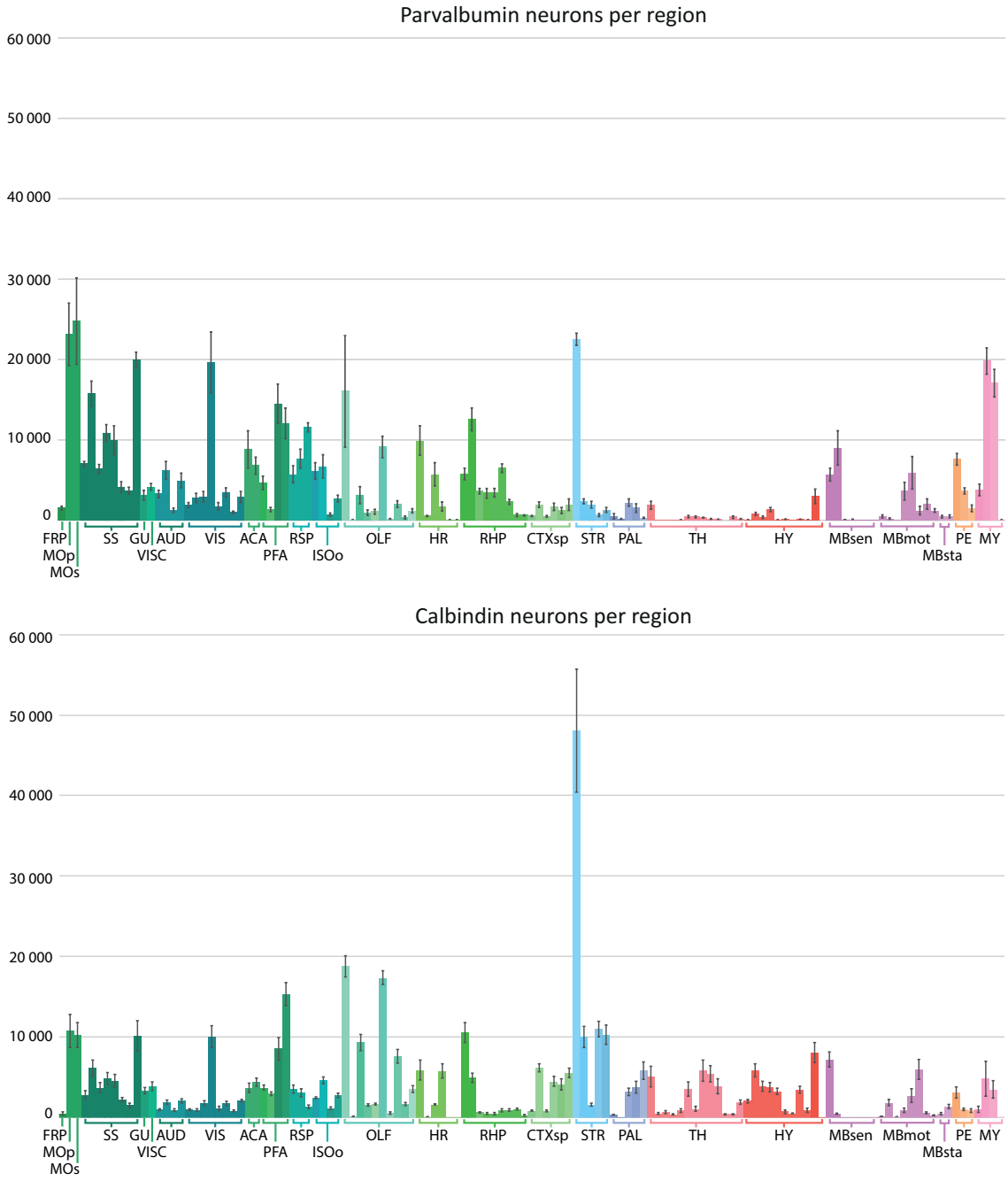

Parvalbumin neurons per region
60 000
50 000
40 000
30 000
20 000
10 000
0
FRP
MOp
MOs
SS
GU
VISC
AUD
VIS
ACA
PFA
RSP
ISOo
OLF
HR
RHP
CTXsp
STR
PAL
TH
HY
MBsen
MBmot
MBsta
PE
MY
Calbindin neurons per region
60 000
50 000
40 000
30 000
20 000
10 000
0
FRP
MOp
MOs
SS
GU
VISC
AUD
VIS
ACA
PFA
RSP
ISOo
OLF
HR
RHP
CTXsp
STR
PAL
TH
HY
MBsen
MBmot
MBsta
PE
MY

Figure 4

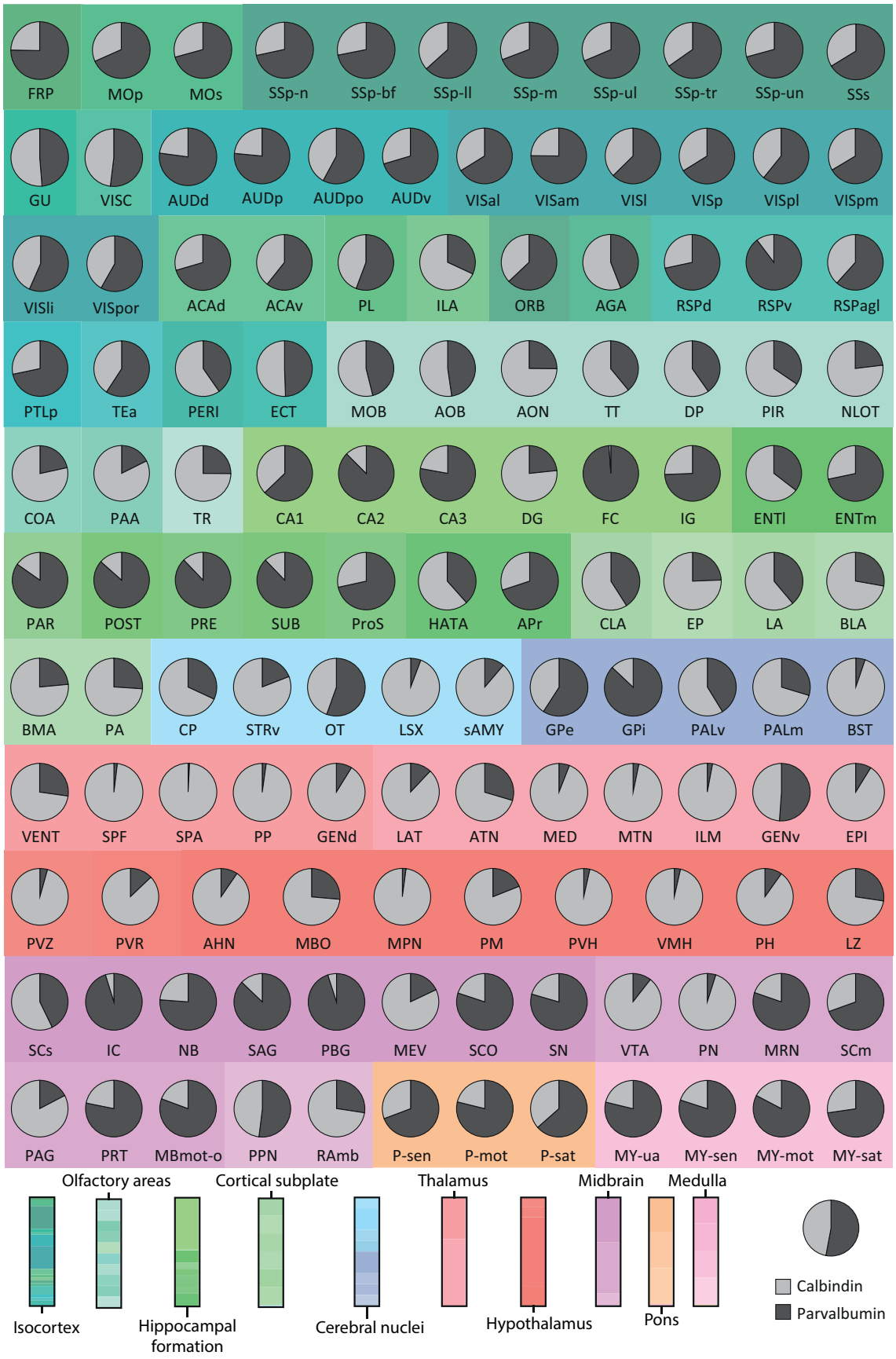

FRP
MOp
MOs
SSp-n
SSp-bf
SSp-ll
SSp-m
SSp-ul
SSp-tr
SSp-un
SSs
GU
VISC
AUDd
AUDp
AUDpo
AUDv
VISal
VISam
VISl
VISp
VISpl
VISpm
VISli
VISpor
ACAd
ACAv
PL
ILA
ORB
AGA
RSPd
RSPv
RSPagl
PTLp
TEa
PERI
ECT
MOB
AOB
AON
TT
DP
PIR
NLOT
COA
PAA
TR
CA1
CA2
CA3
DG
FC
IG
ENTl
ENTm
PAR
POST
PRE
SUB
ProS
HATA
APr
CLA
EP
LA
BLA
BMA
PA
CP
STRv
OT
LSX
sAMY
GPe
GPi
PALv
PALm
BST
VENT
SPF
SPA
PP
GENd
LAT
ATN
MED
MTN
ILM
GENv
EPI
PVZ
PVR
AHN
MBO
MPN
PM
PVH
VMH
PH
LZ
SCs
IC
NB
SAG
PBG
MEV
SCO
SN
VTA
PN
MRN
SCm
PAG
PRT
MBmot-o
PPN
RAmb
P-sen
P-mot
P-sat
MY-ua
MY-sen
MY-mot
MY-sat
Olfactory areas
Cortical subplate
Thalamus
Midbrain
Medulla
Isocortex
Hippocampal formation
Cerebral nuclei
Hypothalamus
Pons
Calbindin
Parvalbumin

Figure 5

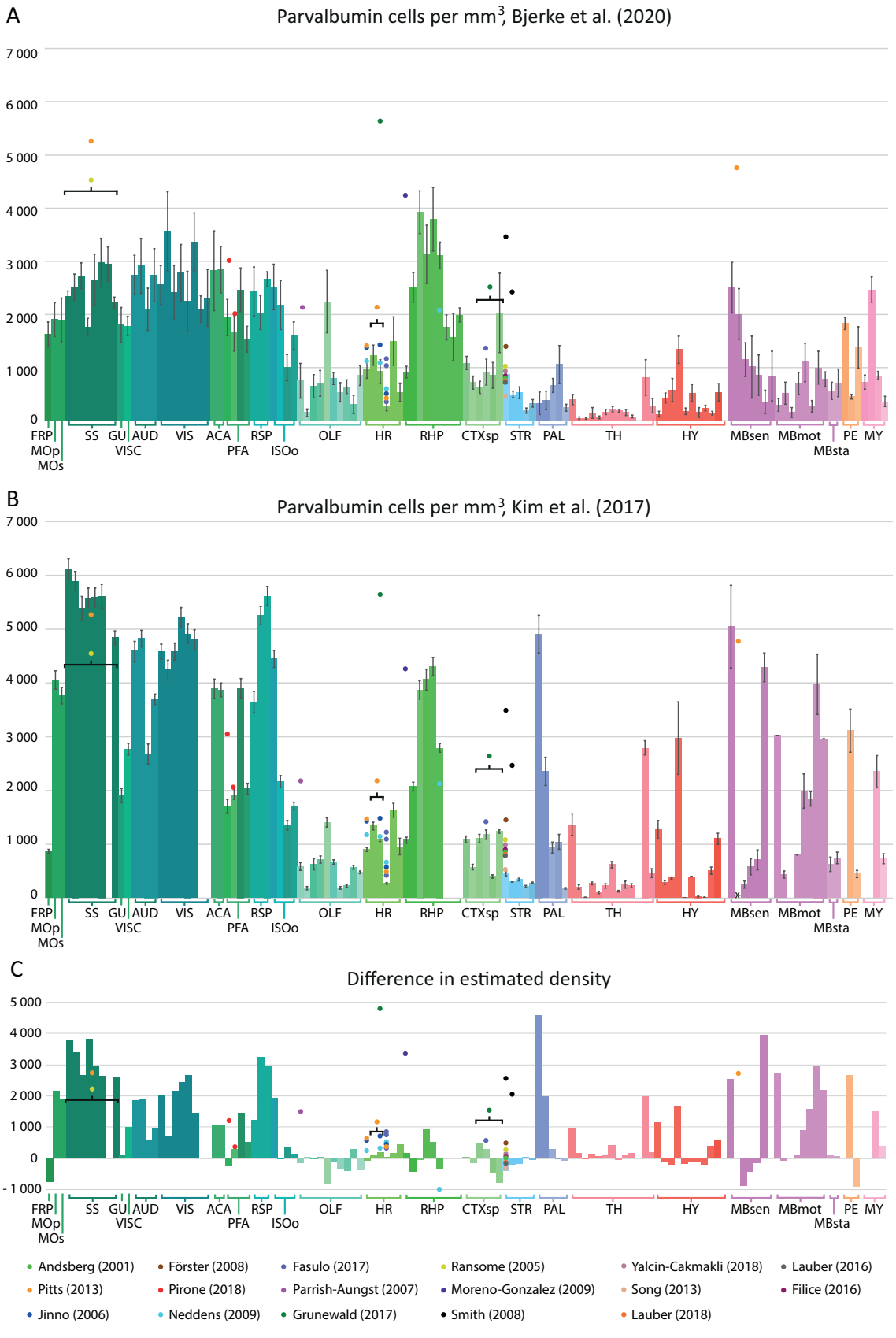

A
Parvalbumin cells per mm3, Bjerke et al. (2020)
7 000
6 000
5 000
4 000
3 000
2 000
1 000
0
FRP
MOp
MOs
SS
GU
VISC
AUD
VIS
ACA
PFA
RSP
ISOo
OLF
HR
RHP
CTXsp
STR
PAL
TH
HY
MBsen
MBmot
MBsta
PE
MY
B
Parvalbumin cells per mm3, Kim et al. (2017)
C
Difference in estimated density
5 000
4 000
3 000
2 000
1 000
0
- 1 000
Andsberg (2001)
Förster (2008)
Fasulo (2017)
Ransome (2005)
Yalcin-Cakmakli (2018)
Lauber (2016)
Pitts (2013)
Pirone (2018)
Parrish-Aungst (2007)
Moreno-Gonzalez (2009)
Song (2013)
Filice (2016)
Jinno (2006)
Neddens (2009)
Grunewald (2017)
Smith (2008)
Lauber (2018)

Figure 6

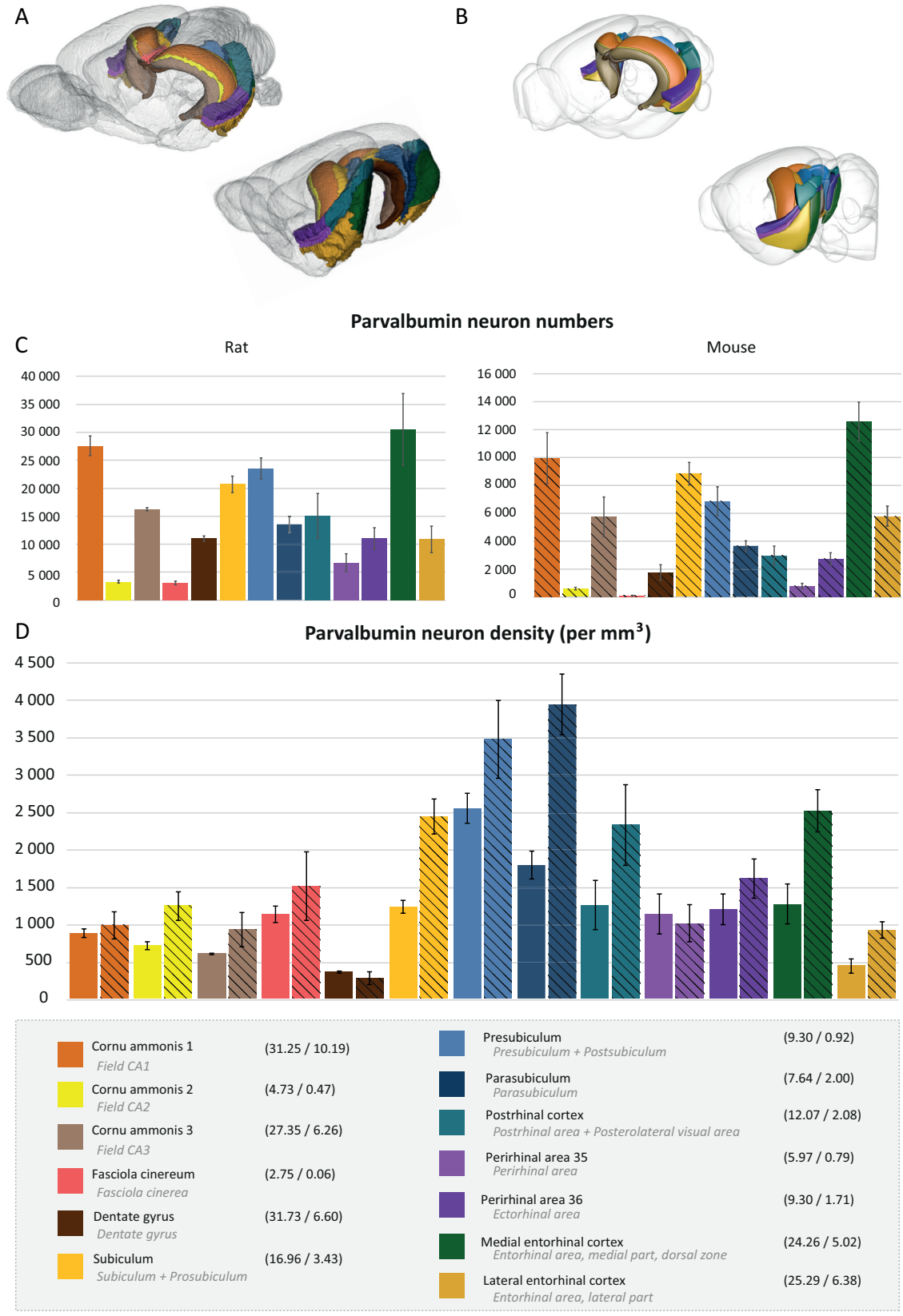

A
B
Parvalbumin neuron numbers
C
Rat
40 000
35 000
30 000
25 000
20 000
15 000
10 000
5 000
0
Mouse
16 000
14 000
12 000
10 000
8 000
6 000
4 000
2 000
0
D
Parvalbumin neuron density (per mm³)
4 500
4 000
3 500
3 000
2 500
2 000
1 500
1 000
500
0
Cornu ammonis 1
Field CA1
(31.25 / 10.19)
Cornu ammonis 2
Field CA2
(4.73 / 0.47)
Cornu ammonis 3
Field CA3
(27.35 / 6.26)
Fasciola cinereum
Fasciola cinerea
(2.75 / 0.06)
Dentate gyrus
Dentate gyrus
(31.73 / 6.60)
Subiculum
Subiculum + Prosubiculum
(16.96 / 3.43)
Presubiculum
Presubiculum + Postsubiculum
(9.30 / 0.92)
Parasubiculum
Parasubiculum
(7.64 / 2.00)
Postrhinal cortex
Postrhinal area + Posterolateral visual area
(12.07 / 2.08)
Perirhinal area 35
Perirhinal area
(5.97 / 0.79)
Perirhinal area 36
Ectorhinal area
(9.30 / 1.71)
Medial entorhinal cortex
Entorhinal area, medial part, dorsal zone
(24.26 / 5.02)
Lateral entorhinal cortex
Entorhinal area, lateral part
(25.29 / 6.38)

Figure 7

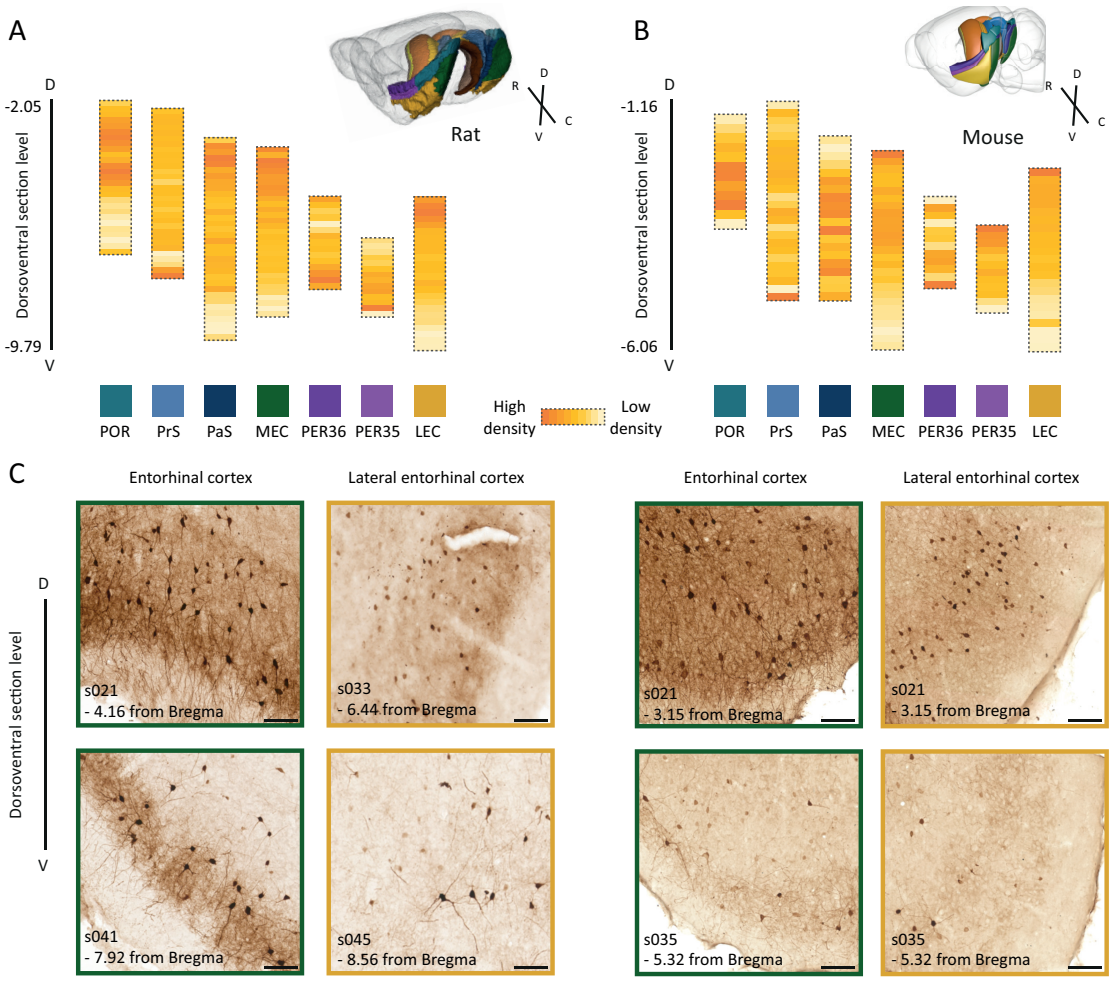
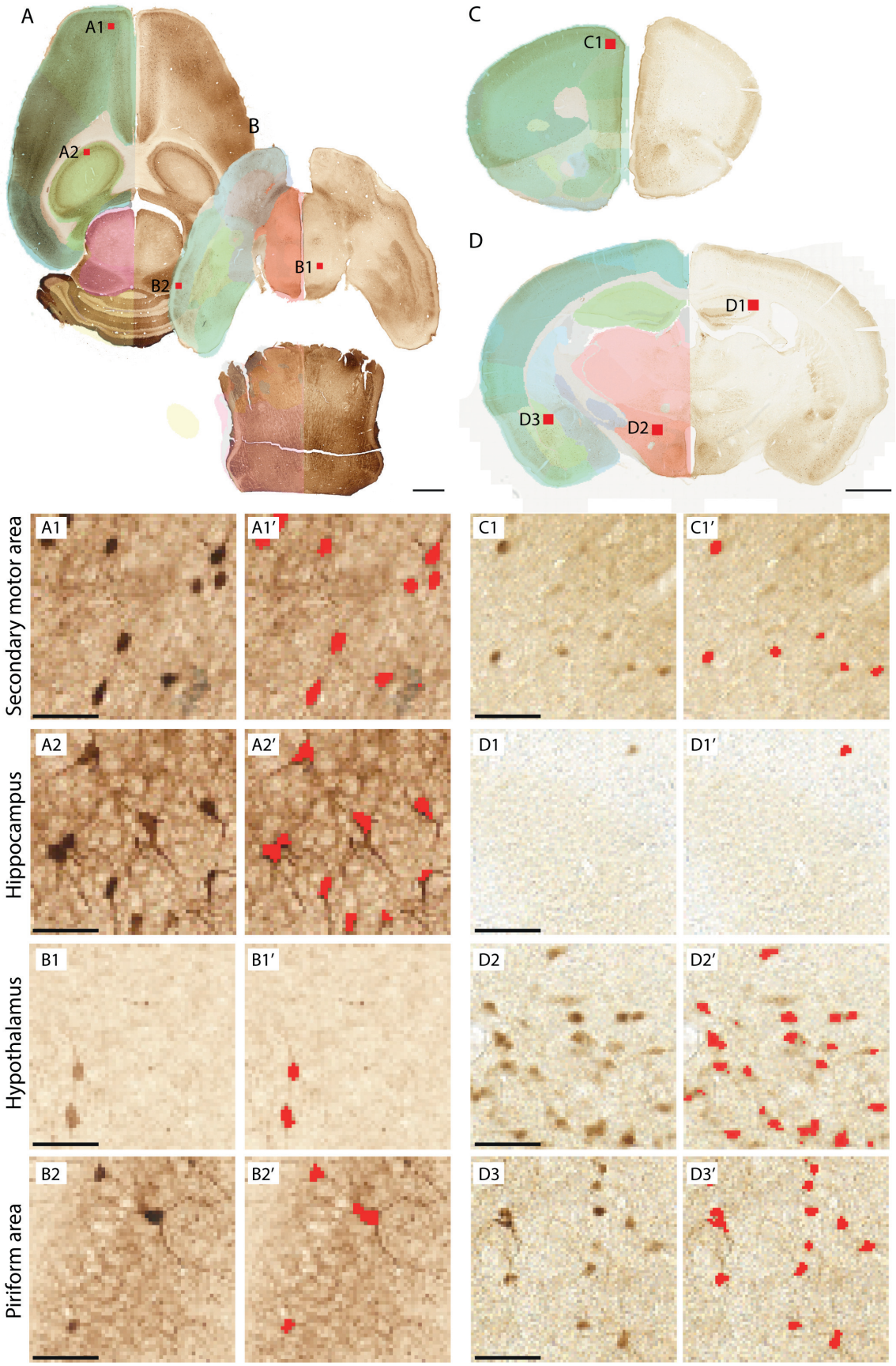

A
B
C
Rat
Mouse
D
V
R
C
-2.05
-9.79
-1.16
-6.06
Dorsoventral section level
POR
PrS
PaS
MEC
PER36
PER35
LEC
High density
Low density
Entorhinal cortex
Lateral entorhinal cortex
s021
- 4.16 from Bregma
s033
- 6.44 from Bregma
s021
- 3.15 from Bregma
s021
- 3.15 from Bregma
s041
- 7.92 from Bregma
s045
- 8.56 from Bregma
s035
- 5.32 from Bregma
s035
- 5.32 from Bregma

Figure 8


A
A1
A2
B
B1
B2
C
C1
D
D1
D2
D3
Secondary motor area
Hippocampus
Hypothalamus
Piriform area
A1
A1′
C1
C1′
A2
A2′
D1
D1′
B1
B1′
D2
D2′
B2
B2′
D3
D3′

Volume: region volume in mm³ **Number:** bilateral total number estimate **Density:** estimated number per mm³		Volume	Parvalbumin		Calbindin	
			Number	Density	Number	Density
Isocortex (ISO)						
FRP	Frontal pole, cerebral cortex	0.97	1 579 ± 223	1622 ± 229	518 ± 214	532 ± 219
MOp	Primary motor area	12.19	23 170 ± 3 895	1 901 ± 319	10 727 ± 2 049	880 ± 168
MOs	Secondary motor area	13.10	24 779 ± 5 368	1 892 ± 410	10 254 ± 1 558	783 ± 119
Somatosensory areas (SS)						
SSp-n	Primary somatosensory area, nose	3.02	7 053 ± 307	2 336 ± 102	2 795 ± 513	926 ± 170
SSP-bfd	Primary somatosensory area, barrel field	6.29	15 777 ± 1 572	2 509 ± 250	6 164 ± 10 05	980 ± 160
SSP-ll	Primary somatosensory area, lower limb	2.35	6 408 ± 574	2 724 ± 244	3 711 ± 609	1 578 ± 259
SSP-m	Primary somatosensory area, mouth	6.21	10 933 ± 1 027	1 761 ± 165	4 854 ± 748	782 ± 120
SSP-ul	Primary somatosensory area, upper limb	3.77	9 988 ± 1 811	2 650 ± 480	4 566 ± 819	1 211 ± 217
SSP-tr	Primary somatosensory area, trunk	1.40	4 167 ± 637	2 973 ± 454	2 218 ± 253	1 582 ± 181
SSP-un	Primary somatosensory area, unassigned	1.26	3 716 ± 410	2 944 ± 325	1 537 ± 277	1 218 ± 220
SSs	Supplemental somatosensory area	9.01	19 948 ± 1 004	2 212 ± 111	10 155 ± 1 908	1 126 ± 212
Gustatory and visceral areas						
GU	Gustatory areas	1.77	3 174 ± 588	1 795 ± 332	3 343 ± 423	1 890 ± 239
VISC	Visceral areas	2.35	4 184 ± 431	1 781 ± 184	3 874 ± 521	1 649 ± 222
Auditory areas (AUD)						
AUDd	Dorsal auditory area	1.21	3 326 ± 454	2 738 ± 373	987 ± 133	813 ± 109
AUDp	Primary auditory area	2.15	6 259 ± 1 115	2 913 ± 519	1 920 ± 287	893 ± 133
AUDpo	Posterior auditory area	0.61	1 283 ± 245	2 096 ± 400	934 ± 136	1 526 ± 222
AUDv	Ventral auditory area	1.82	4 966 ± 911	2 731 ± 501	2 087 ± 271	1 148 ± 149
Visual areas (VIS)						
VISal	Anterolateral visual area	0.76	1 937 ± 275	2 555 ± 362	990 ± 104	1 306 ± 137
VISam	Anteromedial visual area	0.79	2 825 ± 590	3 565 ± 744	933 ± 104	1 178 ± 204
VISl	Lateral visual area	1.23	2 967 ± 619	2 421 ± 505	1 772 ± 308	1 446 ± 251
VISp	Primary visual area	7.07	19 680 ± 3 771	2 782 ± 533	10 063 ± 1 333	1 423 ± 288
VISpl	Posterolateral visual area	0.79	1 786 ± 450	2 248 ± 566	1 150 ± 292	1 447 ± 367
VISpm	Posteromedial visual area	1.04	3 525 ± 572	3 361 ± 546	1 764 ± 285	1 682 ± 272
VISli	Laterointermediate area	0.49	1 032 ± 126	2 097 ± 256	791 ± 121	1 608 ± 246
VISpor	Postrhinal area	1.29	2 965 ± 687	2 312 ± 535	2 131 ± 165	1 662 ± 129
Anterior cingulate areas (ACA)						
ACAd	Anterior cingulate area, dorsal part	3.13	8 835 ± 2 355	2 821 ± 752	3 702 ± 570	1 182 ± 182
ACAv	Anterior cingulate area, ventral part	2.41	6 837 ± 1 059	2 839 ± 440	4 428 ± 505	1 838 ± 210
Prefrontal areas (PFA)						
PL	Prelimbic area	2.41	4 673 ± 827	1 939 ± 343	3 713 ± 322	1 541 ± 133
ILA	Infralimbic area	0.84	1 387 ± 284	1 643 ± 337	2 960 ± 211	3 506 ± 249
ORB	Orbital area	5.90	14 517 ± 2 448	2 460 ± 415	8 547 ± 1 376	1 448 ± 233
AGA	Agranular insular area	7.89	12 079 ± 1 935	1 530 ± 245	15 332 ± 1 448	1 942 ± 183
Retrosplenial areas (RSP)						
RSPagl	Retrosplenial area, lateral agranular part	2.35	5 733 ± 1 084	2 432 ± 460	3 563 ± 515	1 512 ± 218
RSPd	Retrosplenial area, dorsal part	3.80	7 713 ± 1 201	2 030 ± 316	3 052 ± 494	803 ± 130
RSPv	Retrosplenial area, ventral part	4.35	11 590 ± 597	2 666 ± 137	1 331 ± 239	306 ± 55
Other isocortical areas (ISOo)						
PTLp	Posterior parietal association areas	2.46	6 181 ± 1 057	2 516 ± 430	2 439 ± 94	993 ± 38
TEa	Temporal association areas	3.10	6 741 ± 1 436	2 173 ± 463	4 645 ± 425	1 497 ± 137
PERI	Perirhinal area	0.79	789 ± 195	999 ± 247	1 169 ± 169	1 481 ± 214
ECT	Ectorhinal area	1.71	2 735 ± 447	1 597 ± 261	2 795 ± 277	1 632 ± 162
Olfactory areas (OLF)						
MOB	Main olfactory bulb	21.49	16 071 ± 6 949	748 ± 323	18 784 ± 1 355	874 ± 63
AOB	Accessory olfactory bulb	0.65	98 ± 47	150 ± 72	107 ± 31	164 ± 48
AON	Accessory olfactory nucleus	4.87	3 139 ± 1 056	644 ± 217	9 332 ± 987	1 916 ± 203
TT	Taenia tecta	1.44	1 008 ± 353	699 ± 245	1 574 ± 144	1 092 ± 100

DP	Dorsal peduncular nucleus	0.50	1 130 ± 298	2 240 ± 590	1 677 ± 79	3 324 ± 156
PIR	Piriform area	11.57	9 162 ± 1 349	792 ± 117	17 379 ± 865	1 502 ± 75
NLOT	Nucleus of the lateral olfactory tract	0.32	169 ± 59	529 ± 183	563 ± 132	1 763 ± 414
COA	Cortical amygdalar area	3.27	2 089 ± 422	640 ± 129	7 589 ± 843	2 324 ± 258
PAA	Piriform-amygdalar area	1.19	356 ± 207	298 ± 174	1 660 ± 199	1 393 ± 167
TR	Postpiriform transition area	1.40	1 193 ± 268	849 ± 191	3 559 ± 417	2 534 ± 297
Hippocampal formation (HPF)						
Hippocampal region (HR)						
CA1	Field CA1	10.19	9 947 ± 1 832	976 ± 180	5 893 ± 1 257	578 ± 123
CA2	Field CA2	0.48	587 ± 91	1 234 ± 191	84 ± 19	177 ± 40
CA3	Field CA3	6.26	5 752 ± 1 424	918 ± 227	1 638 ± 110	262 ± 17
DG	Dentate gyrus	6.60	1 759 ± 558	267 ± 85	5 778 ± 896	876 ± 136
FC	Fasciola cinerea	0.06	92 ± 28	1 495 ± 457	1 ± 1	16 ± 16
IG	Induseum griseum	0.12	67 ± 22	533 ± 173	23 ± 7	183 ± 54
Retrohippocampal region (RHP)						
ENTl	Entorhinal area, lateral part	6.38	5 804 ± 717	910 ± 112	10 580 ± 1 257	1 659 ± 197
ENTm	Entorhinal area, medial part	5.02	12 576 ± 1 424	2 506 ± 281	4 962 ± 586	989 ± 117
PAR	Parasubiculum	0.93	3 658 ± 377	3 921 ± 404	658 ± 83	706 ± 89
POST	Postsubiculum	1.08	3 396 ± 595	3 132 ± 549	528 ± 112	487 ± 104
PRE	Presubiculum	0.92	3 472 ± 550	3 787 ± 600	479 ± 134	523 ± 146
SUB	Subiculum	2.10	6 521 ± 532	3 108 ± 254	909 ± 221	433 ± 106
ProS	Prosubiculum	1.33	2 326 ± 308	1 751 ± 232	924 ± 141	696 ± 106
HATA	Hippocampo-amygdalar transition area	0.42	662 ± 187	1 571 ± 444	1 061 ± 128	2 519 ± 303
APr	Area prostriata	0.32	649 ± 44	1 986 ± 136	279 ± 64	853 ± 263
Cortical subplate (CTXsp)						
CLA	Claustrum	0.55	596 ± 68	1 082 ± 123	859 ± 64	1 559 ± 116
EP	Endopiriform nucleus	2.79	1 994 ± 345	714 ± 123	6 186 ± 477	2 215 ± 171
LA	Lateral amygdalar nucleus	0.84	527 ± 98	627 ± 116	829 ± 130	986 ± 154
BLA	Basolateral amygdalar nucleus	1.90	1 726 ± 462	909 ± 243	4 486 ± 631	2 361 ± 332
BMA	Basomedial amygdalar nucleus	1.49	1 262 ± 367	850 ± 247	4 109 ± 681	2 766 ± 458
PA	Posterior amygdalar nucleus	0.97	1 956 ± 723	2 023 ± 748	5 554 ± 569	5 745 ± 589
Cerebral nuclei (CNU)						
Striatum (STR)						
CP	Caudoputamen	26.01	22 523 ± 778	866 ± 30	48 064 ± 7 665	1 847 ± 295
STRv	Striatum ventral region	4.83	2 367 ± 309	490 ± 64	10 027 ± 1 326	2 076 ± 275
OT	Olfactory tubercle	3.82	1 990 ± 424	520 ± 111	1 597 ± 182	418 ± 48
LSX	Lateral septal complex	3.56	666 ± 178	187 ± 50	11 005 ± 982	3 088 ± 276
sAMY	Striatum-like amygdalar nuclei	4.05	1 307 ± 301	323 ± 74	10 291 ± 1 188	2 540 ± 293
Pallidum (PAL)						
GPe	Globus pallidus, external segment	1.56	506 ± 330	323 ± 211	350 ± 38	223 ± 24
GPi	Globus pallidus, internal segment	0.42	161 ± 73	381 ± 172	24 ± 7	57 ± 16
PALv	Pallidum, ventral region	3.39	2 243 ± 447	662 ± 132	3 196 ± 462	942 ± 136
PALm	Pallidum, medial region	1.51	1 589 ± 539	1 052 ± 357	3 766 ± 712	2 492 ± 471
BST	Bed nuclei of the stria terminalis	1.34	324 ± 88	241 ± 65	5 832 ± 1 067	4 349 ± 796
Thalamus (TH)						
VENT	Ventral group of the dorsal thalamus	4.86	1 900 ± 473	391 ± 97	5 078 ± 1 282	1 045 ± 264
SPF	Subparafascicular nucleus	0.20	10 ± 5	48 ± 22	507 ± 97	2 474 ± 474
SPA	Subparafascicular area	0.13	5 ± 3	41 ± 21	658 ± 202	4 964 ± 1 526
PP	Peripeduncular area	0.06	9 ± 6	144 ± 102	389 ± 91	6 359 ± 1 482
GENd	Geniculate group, dorsal thalamus	1.42	87 ± 39	61 ± 28	893 ± 241	628 ± 170
LAT	Lateral group of the dorsal thalamus	3.08	484 ± 146	157 ± 48	3 540 ± 860	1 151 ± 279
ATN	Anterior group of the dorsal thalamus	2.16	455 ± 105	211 ± 49	1 084 ± 314	502 ± 146
MED	Medial group of the dorsal thalamus	2.08	371 ± 49	184 ± 24	5 844 ± 1 328	2 897 ± 658
MTN	Midline group of the dorsal thalamus	1.19	181 ± 66	153 ± 56	5 438 ± 976	4 585 ± 823
ILM	Intralaminar nuclei of the dorsal thalamus	1.60	117 ± 43	73 ± 27	3 880 ± 932	2 424 ± 582
RT	Reticular nucleus of the thalamus	-	-	-	416 ± 70	287 ± 48
GENv	Geniculate group, ventral thalamus	0.52	424 ± 177	809 ± 338	405 ± 71	773 ± 136
EPI	Epithalamus	0.66	185 ± 87	279 ± 131	1 920 ± 285	2 888 ± 429
Hypothalamus (HY)						
PVZ	Periventricular zone	0.78	93 ± 42	119 ± 53	2 046 ± 240	2 614 ± 307
PVR	Periventricular region	2.04	869 ± 189	425 ± 92	5 859 ± 815	2 864 ± 398
AHN	Anterior hypothalamic nucleus	0.71	410 ± 155	578 ± 218	3 881 ± 592	5 462 ± 833
MBO	Mammillary body	1.02	1 353 ± 263	1 332 ± 258	3 779 ± 578	3 720 ± 569
MPN	Medial preoptic nucleus	0.41	70 ± 25	173 ± 61	3 253 ± 396	8 003 ± 975

PM	Premammillary nuclei	0.33	169 ± 55	517 ± 169	721 ±234	2 202 ± 716
PVH	Paraventricular hypothalamic nucleus, descending division	0.13	20 ± 13	151 ± 97	542 ± 29	4 150 ± 219
VMH	Ventromedial hypothalamic nucleus	0.55	127 ± 29	232 ± 52	3 460 ± 404	6 298 ± 736
PH	Posterior hypothalamic nucleus	0.71	98 ± 27	138 ± 39	893 ± 255	1 263 ± 360
LZ	Hypothalamic lateral zone	5.69	3 026 ± 930	532 ± 164	8 049 ± 1 250	1 416 ± 220
ME	Median eminence	0.09	0 ± 0	0 ± 0	6 ± 5	74 ± 60
Midbrain (MB)						
Midbrain, sensory related (MBsen)						
SCs	Superior colliculus, sensory related	2.16	5 396 ± 1 031	2 502 ± 478	7 213 ± 928	3 345 ± 430
IC	Inferior colliculus	4.44	8 911 ± 2 131	2 004 ± 479	450 ± 65	101 ± 15
NB	Nucleus of the brachium of the inferior colliculus	0.09	101 ± 28	1 148 ± 320	32 ± 11	359 ± 129
SAG	Nucleus sagulum	0.10	99 ± 58	1 006 ± 586	15 ± 4	148 ± 43
PBG	Parabigeminal nucleus	0.04	38 ± 17	858 ± 381	3 ± 2	45 ± 45
MEV	Midbrain trigeminal nucleus	0.01	3 ± 2	347 ± 222	15 ± 3	1 585 ± 289
SCO	Subcommissural organ	0.01	12 ± 7	834 ± 469	3 ± 2	208 ± 138
SN	Substantia nigra	1.74	513 ± 208	293 ± 119	132 ± 39	76 ± 23
Midbrain, motor related (MBmot)						
VTA	Ventral tegmental area	0.43	220 ± 90	512 ± 210	1 860 ± 392	4 339 ± 914
PN	Paranigral nucleus	0.02	3 ± 2	151 ± 97	64 ± 13	2 951 ± 620
MRN	Midbrain reticular nucleus	5.28	3 674 ± 1 099	696 ± 208	901 ± 241	171 ± 46
SCm	Superior colliculus, motor related	5.65	6 189 ± 2 048	1 095 ± 362	2 716 ± 836	480 ± 148
PAG	Periaqueductal grey	4.69	1 250 ± 528	266 ± 112	5 962 ± 1 231	1 270 ± 262
PRT	Pretectal region	2.08	2 070 ± 651	995 ± 313	578 ± 119	278 ± 57
MBm-o	Midbrain, motor related, other	1.60	1 233 ± 228	775 ± 144	296 ± 55	186 ± 35
Midbrain, behavioral state related (MBsta)						
PPN	Pedunculopontine nucleus	0.89	498 ± 147	558 ± 165	461 ± 133	516 ± 149
RAmb	Midbrain raphe nuclei	0.73	518 ± 190	706 ± 259	1 373 ± 272	1 872 ± 371
Pons (PE)						
P-sen	Pons, sensory related	3.85	7 052 ± 444	1 832 ± 115	3 128 ± 691	812 ± 179
P-mot	Pons, motor related	8.16	3 668 ± 360	450 ± 44	985 ± 156	121 ± 19
P-sat	Pons, behavioral state related	1.10	1 515 ± 428	1 378 ± 389	871 ± 248	792 ± 226
Medulla (MY)						
MY-ua	Medulla, unassigned	5.32	3 835 ± 724	721 ± 136	1 032 ± 409	194 ± 77
MY-sen	Medulla, sensory related	7.83	19 301 ± 1 879	2 465 ± 240	4 821 ± 2 155	616 ± 275
MY-mot	Medulla, motor related	19.75	16 682 ± 1 576	845 ± 80	3 493 ± 1 090	177 ± 55
MY-sat	Medulla, behavioral state related	0.22	80 ± 22	358 ± 97	30 ± 10	134 ± 46

Validity and reliability of QUINT procedures for cell counting

Methods

Validation of segmentations

The results obtained with the QUINT workflow were benchmarked against prior manual counts from both parvalbumin and calbindin stained material that were used as a guide during the segmentation process. Manual counting was performed on a subset of the material by annotating each labelled cell in Adobe Photoshop (RRID:SCR_014199). The annotated cells were subsequently quantified using Nutil, as described above. Manual counting was performed for each section throughout the calbindin stained anterior cingulate cortex, the parvalbumin stained entorhinal cortex, and for one whole hemisphere from each of the stains. The regions of interest were defined by the QuickNII atlas maps for the evaluated sections and were thus identical for the quantification of manual and ilastik segmentations.

Reliability of segmentations

We assessed both intra- and interrater reliability by segmenting material from one of the series (parvalbumin stained sections from rat 25205) several times. Intrarater reliability was assessed by one researcher segmenting the same material three times. The first and second segmentation was performed within the same week, while the last one was done six weeks later.

In a pilot study of interrater reliability, we considered the reliability of segmentations obtained by three researchers using the same material, but without clear guidelines for what to consider a labelled cell. We observed, both qualitatively and quantitatively, that results varied considerably between the three researchers. We therefore set out to do a more systematic assessment of interrater reliability, with clear instructions on how to perform the segmentation with ilastik. One researcher performed an initial segmentation and wrote instructions for its replication. The same materials

and instructions were then presented to five other researchers for segmentation. The instructions gave a description of objects considered to be cells and example images of segmentation results. In addition, information was given on how to train the ilastik classifier, with instruction to not label pixels in oversaturated areas (specifically, the cerebellum and reticular nucleus of thalamus). Researchers were also given access to the segmentations overlaid on the original (high-resolution) images via an online viewer system, and were instructed to use these actively to verify similarity of segmentations. The full instructions as given to the researchers is included in Supplementary File 3. Each researcher then ran their classifier on the full set of images from rat 25205. Objects in the resulting segmented images were quantified using Nutil quantifier, as described above.

Results

Validity of QUINT results

Parvalbumin-stained material. We compared the numbers of parvalbumin positive cells obtained by the QUINT workflow with manual counts. We here first present the results for the counts made from one hemisphere and go on to elaborate the results throughout the entorhinal cortex.

We quantified cells in one hemisphere (section 028 from rat 25205) using both manual and ilastik segmentation. The difference between counts obtained by the two approaches was 9 % in the neocortex (manual count = 2175, QUINT count = 2379), 3 % in the hippocampal and parahippocampal regions (manual count = 540, QUINT count = 555), and 15 % for striatal and pallidal regions (manual count = 295, QUINT count = 340).

Comparison of QUINT results with manual counts of parvalbumin neurons from throughout the entorhinal cortex (from one hemisphere, sections s019 through s044) showed that the classifier generated numbers very similar to the manual approach. The total number of PV positive neurons counted were 3400 with manual counts and 3560 with the QUINT approach (5 % difference). To

obtain total number estimates for the whole region, we used Abercrombie's formula to correct for double-counting and multiplied by the section interval. With this approach, the total number estimates in the unilateral medial entorhinal cortex were 14 836 by use of the manual counts and 15 535 by use of QUINT counts.

The results from the entorhinal cortex per section varied from 0-16 % between the two approaches (on average 3 % difference), and did not seem to relate to whether or not the section was included in the training material (on average 5 % difference in training sections and 6 % difference in non-included sections). The results from both methods per section are summarised in Fig X. Note that the last section that included the medial entorhinal cortex (s044) was excluded from the section-by-section analysis because it only contained three manually counted cells. By qualitative inspection of each section with both manual and ilastik segmentations overlaid, we observed that the differences were mainly related to objects that were hard to identify unambiguously. Some of the differences might also be due to the object splitting feature in the Nutil software, which causes objects placed on the border between regions to be randomly assigned to one of them. Because larger objects will be more likely to intersect region borders, this feature affected ilastik segmentations more than the manual ones. Cells in the latter were represented by only a single pixel (see lower panel in Figure 1) and thus could not intersect two regions, while whole cells were segmented with ilastik and relatively more often had pixels in two different regions.

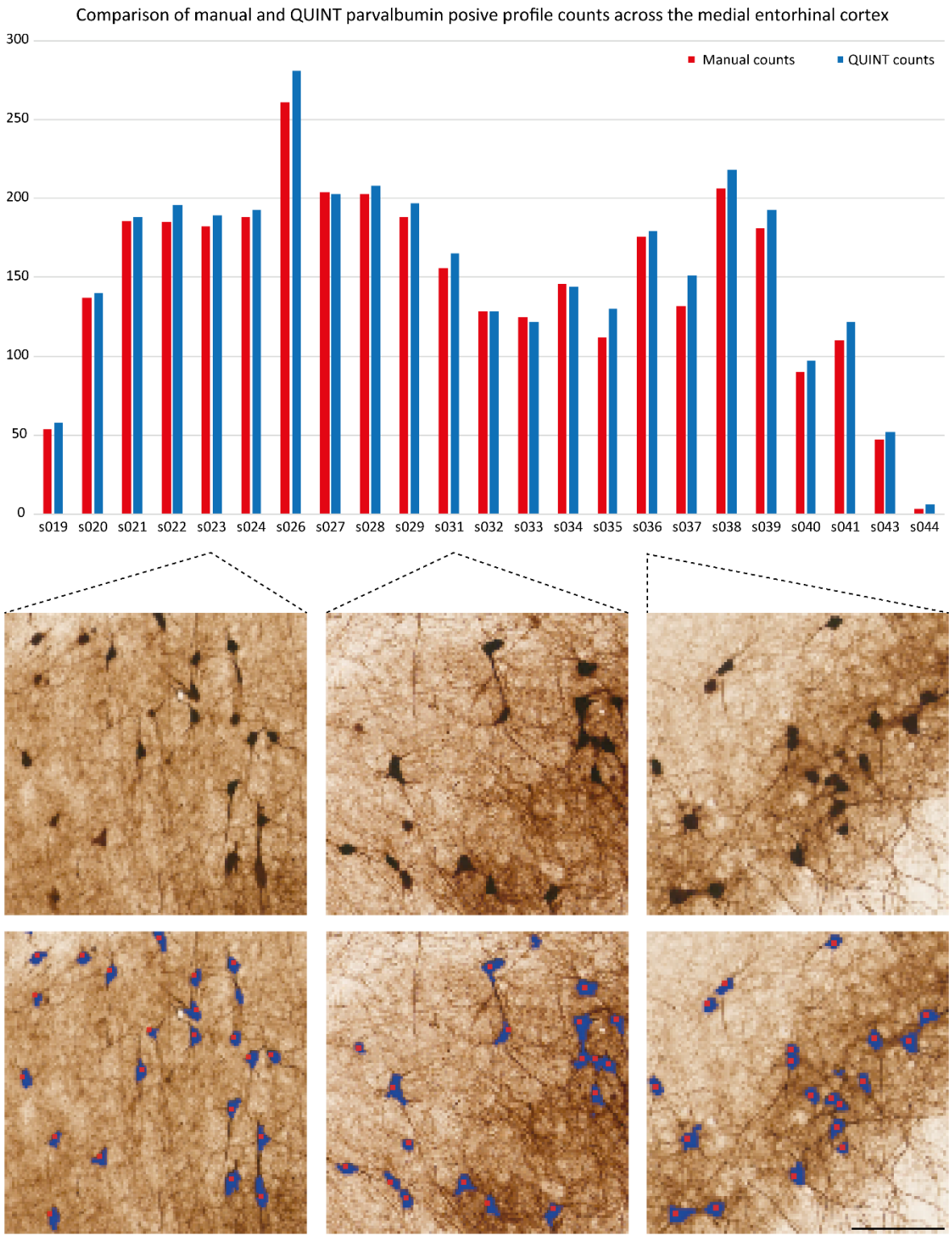


Figure 1. Comparison of manual and QUINT counts of parvalbumin positive profile counts across the medial entorhinal cortex. The graph shows manual counts (red bars) and QUINT

counts (blue bars) from horizontal sections throughout the medial entorhinal cortex. Example images showing the manual and ilastik segmentations overlayed to the parvalbumin stained material are shown in the lower panel. Scale bar: 100 µm.

Calbindin-stained material. Numbers of calbindin positive cells obtained using the QUINT workflow were compared manual counts. We here first present the results for the counts made from one hemisphere, and go on to elaborate the results throughout the anterior cingulate cortex.

Numbers of cells obtained by the QUINT workflow were compared to manual counts from one hemisphere in section 151 from mouse 10. The difference between the results obtained by the two methods was 10 % in the cortex (manual count = 654, QUINT count = 719), - 2 % in the hypothalamus (manual count = 98, QUINT count = 96), and 10 % in striatal and pallidal regions (manual count = 1371, QUINT count = 1513). A more pronounced difference was seen in olfactory regions (44 %, manual count = 63, QUINT count = 91). By inspecting manual and QUINT segmentations, we observed that most differences were due to the very lightly stained profiles typically present in layer II / III of cortical regions. These are ambiguous profiles that are difficult to classify regardless of method, but are probably underrepresented by our classifier, as discussed in the main publication.

We furthermore compared QUINT counts with manual ones throughout the anterior cingulate cortex. The difference between the two methods for this region was 16 % (manual count = 905 cells, QUINT count = 961 cells). Again, qualitative inspection of the segmentation images resulting from the two methods showed that differences were mainly related to lightly stained cells. These are overrepresented in layer II / III of cortical regions, and caution is therefore warranted when considering data from these areas specifically. As noted above, the QUINT results are in accordance with manual counts when considering the cortex as a whole.

Reliability of QUINT results

Intrarater reliability. To assess the reliability of the segmentation process given the same researcher and the same material, one researcher did the segmentation three times. Variability was very low in the hippocampus and parahippocampal areas (1 - 3 % difference) and neocortex (0.2 - 2 % difference), relatively low in the striatum (3 – 10 % difference), globus pallidus (1 – 8 % difference), and hypothalamus (10 – 11 % difference), and slightly higher in the substantia nigra (8 – 14 % difference) and basal forebrain (3 – 13 % difference). Results from the intrarater reliability test for all regions are summarised in Figure 2, and all the derived data are included in Supplementary file 4. The variability did not appear higher six weeks after making the initial segmentation.

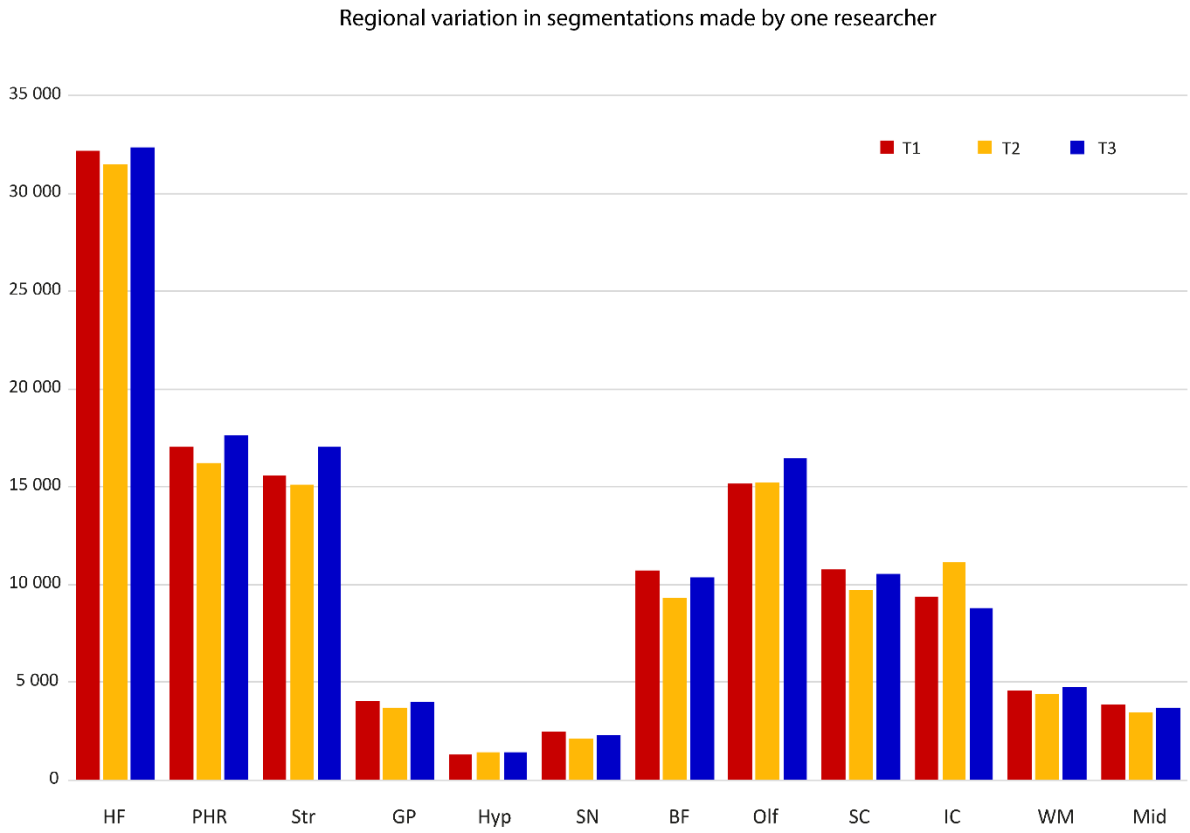


Figure 2. Reliability of segmentations from one researcher. Bar chart showing the results from quantification of segmentations made by one researcher at three different time points (T1-3). Abbreviations: BF, basal forebrain; GP, globus pallidus; HF, hippocampal formation; Hyp, hypothalamus; IC, inferior colliculus; PHR, parahippocampal region; Mid, other midbrain regions Str, striatum; SN, substantia nigra; Olf, olfactory regions; SC, superior colliculus; WM, white matter. A list of which regions of the Waxholm Space rat brain atlas are included in these abbreviations is included in Supplementary file 4.

Interrater reliability. We assessed the reliability of the segmentation results by presenting five researchers with the same material and the same instructions for segmentation. All segmentation results were then combined with the same atlas maps and submitted to Nutil Quantifier for analysis. When summarized according to major brain regions in the atlas, the average difference from the segmentation to be replicated ranged from 8 – 29 % for the different researchers. The quantitative and qualitative results of the reliability analysis is summarized in Figure 3. Interestingly, the average difference per region was relatively low for the researchers who had limited or no experience with the ilastik software (8 – 13 %), but considerably higher for those that had experience (27 and 29 %). The derived data for the interrater reliability analysis (with counts per all brain regions according to the Waxholm Space atlas of the rat brain) is included in Supplementary File 4.

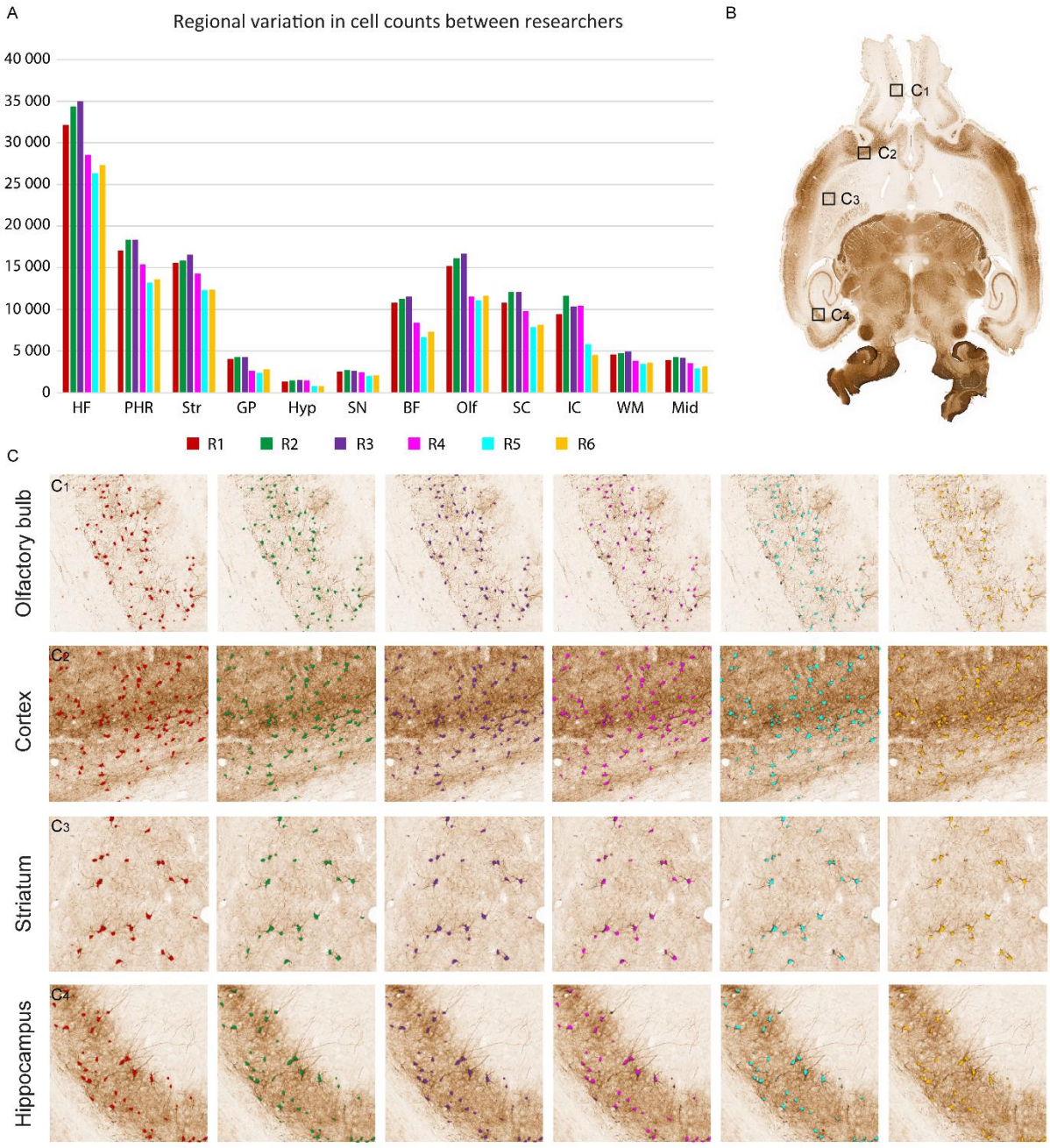


Figure 3. Reliability of segmentations across researchers. (A) Bar chart showing variations between different researchers segmentation results. One researcher (R1, red bars) performed a segmentation and provided documentation for replication of the analysis by researchers 2-5. Examples from the segmentations produced by each researcher is shown as overlays on the original images in the lower panel (C1-C4, colour coded in accordance with the bar graph). Abbreviations: BF, basal forebrain; GP, globus pallidus; HF, hippocampal formation; Hyp, hypothalamus; IC, inferior colliculus; PHR, parahippocampal region; Mid, other midbrain regions Str, striatum; SN, substantia nigra; Olf, olfactory regions; SC, superior colliculus; WM, white matter. A list of which

regions of the Waxholm Space rat brain atlas are included in these abbreviations is included in Supplementary file 4.

The variation of the segmentations made by each researcher systematically deviated from the one to be reproduced (i.e. consistently over- or under-estimated the number of objects), with the exception of two regions from one of the researchers. Qualitatively, we observed that the segmentations that underestimated objects relative to the original segmentation often did not extract the most "extreme" cases of labelling, i.e. very darkly stained cells, or light-to-medium stained cells. The documentation instructed not to segment very light cells, which might have left room for interpretation in such cases.

During ilastik segmentation with the pixel classification workflow, the researcher places labels to train the classifier. We exported and quantified the labels made by each researcher by the default custom regions supported for the WHS by Nutil Quantifier. The classifier to be recreated contained 90 objects labelled as "cell". Among the researchers attempting to recreate it, the number of cells labelled ranged from 33-63. All researchers had placed cell labels in the cortex and hippocampus, and all but one placed labels in olfactory and striatal/pallidal regions as well. For the "background" label, we considered label load instead of number; here, all researchers placed labels across the cortex, fiber tracts and hippocampus, and most placed labels in olfactory regions as well.

Discussion

We show that the profile counts obtained with the QUINT workflow are in accordance with manual counts from the same areas. To convert profile counts into cell number estimates from entire regions, it is necessary to address several sources of bias (Attili et al., 2019). First, when the brain is cut into sections, cells may be split and can therefore appear in more than one section – this leads to a tendency for overcounting. We corrected for this by using Abercrombie's formula (Abercrombie, 1946). Secondly, cells that are located at the border between two atlas regions must

only be counted one, and in these cases Nutil Quantifier assigns the object to only one region. Third, the bias of lost caps (i.e. cell fragments at the edges of a section may be "lost" or invisible; Hedreen, 1998) are inherent to profile counts in histological material. Correcting for this would require estimating a factor based on the observed number of profiles and the true number of cells, the latter which could not be derived from section images. Lastly, cells located in deeper parts of a thick section may be occluded by those in the upper layer, but correcting for this would require dividing the section into layers along the z axis, which also would not be possible in section images. Although it does not address these last two sources of bias, it has been shown that profile counting with Abercrombie's correction yield similar results to both stereology and three-dimensional reconstruction of entire cell populations (Baquet et al., 2009). Thus, to the extent to which our QUINT profile counts accurately reflect manual profile counts, we also consider them to reflect cell counts when corrected and multiplied to represent whole regions.

Our reliability testing indicated that detailed instructions on the criteria for segmentation, with visualisation of the expected outcome, can be effective in allowing researchers to recreate an analysis. In this test, only one parameter (the ilastik segmentations) was different between the researchers. In contrast, reproducibility of scientific findings is typically evaluated across scientific papers where the sources of variability may be many and hard to assess. Although methods sections in scientific reports are intended to provide the sufficient and necessary details to reproduce results, they often lack critical information needed to interpret analyses (Bjerke et al., 2018; Keller et al., 2018). Our instructions were formulated as a step-wise procedure, and may not be representative of a typical methods section, but we believe it can give clues as to the details that are important for researchers to interpret and recreate an analysis. For counting data, including several visual examples of what is considered an object should be considered a minimum. Ideally, some form of

representations of counted objects across the entire quantified material (such as the segmentation images provided here) allows other researchers to gain a deeper understanding of the analytic results.

References

Abercrombie, M. (1946). Estimation of nuclear population from microtome sections. Anat. Rec. *94*, 239–247.

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

Baquet, Z., Williams, D., Brody, J., and Smeyne, R. (2009). A comparison of model-based (2D) and design-based (3D) stereological methods for estimating cell number in the substantia nigra pars compacta (SNpc) of the C57BL/6J mouse. Neuroscience *161*, 1082–1090.

Bjerke, I., Øvsthus, M., Andersson, K., Blixhavn, C., Kleven, H., Yates, S., Puchades, M., Bjaalie, J., and Leergaard, T. (2018). Navigating the murine brain: Toward best practices for determining and documenting neuroanatomical locations in experimental studies. Front. Neuroanat. *12*, 82.

Hedreen, J.C. (1998). Lost caps in histological counting methods. Anat. Rec. *250*, 366–372.

Keller, D., Erö, C., and Markram, H. (2018). Cell densities in the mouse brain: A systematic review. Front. Neuroanat. *12*, 83.

Documentation for segmentation of parvalbumin cells in rat 25205

Goal. The goal of this exercise is to reproduce a segmentation ("target segmentation") made by one researcher as faithfully as possible. The original segmentation was made with the aim of extracting parvalbumin positive cell bodies from DAB-stained images. The cells are generally easy to recognize, but there are variations across the material that makes some profiles more difficult to determine. In the more difficult cases, researchers might have different opinions on what should be considered a cell. I have aimed to provide the sufficient and necessary documentation with respect to what I have considered a cell when segmenting this material. It is important that you consider the provided examples and attempt to replicate this, regardless of whether you agree with the definitions and segmentations made here.

Training images. The training material includes very fifth section throughout the series, starting with s003 and ending with s053. Images are 15 % of original tiff resolution (with original tiffs being 50 % of original CZI file).

Cell criteria. A combined evaluation of the following criteria was used to decide whether an object is a cell or not:

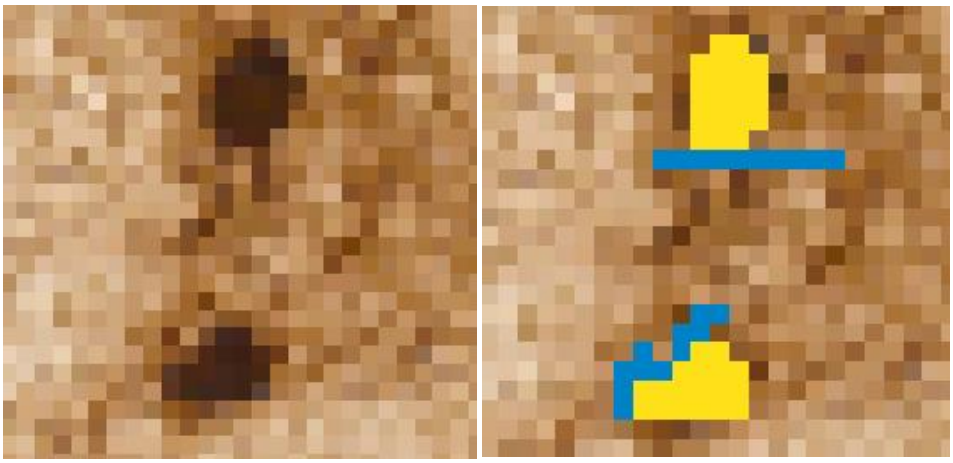
- clearly distinguished from the background
- round, ovoid, triangular, or multipolar in shape
- medium, darkly, or intensely stained
- size substantial enough to support that the object is a cell and not axonal or dendritic fragments

See "Examples of manually annotated cells" below for examples.

Training in ilastik

Certain areas are so heavily labelled or contains so much stained neuropil that they should be excluded from the brain-wide analysis so as not to compromise the quality of the remaining classifier. For this reason, **do not** segment pixels in the cerebellum or reticular nucleus of the thalamus. Before starting, see the examples of manual annotations and target segmentations in the end of this document. High-resolution images with target segmentations overlaid are available via the Navigator system, and can be used as a reference during segmentation.

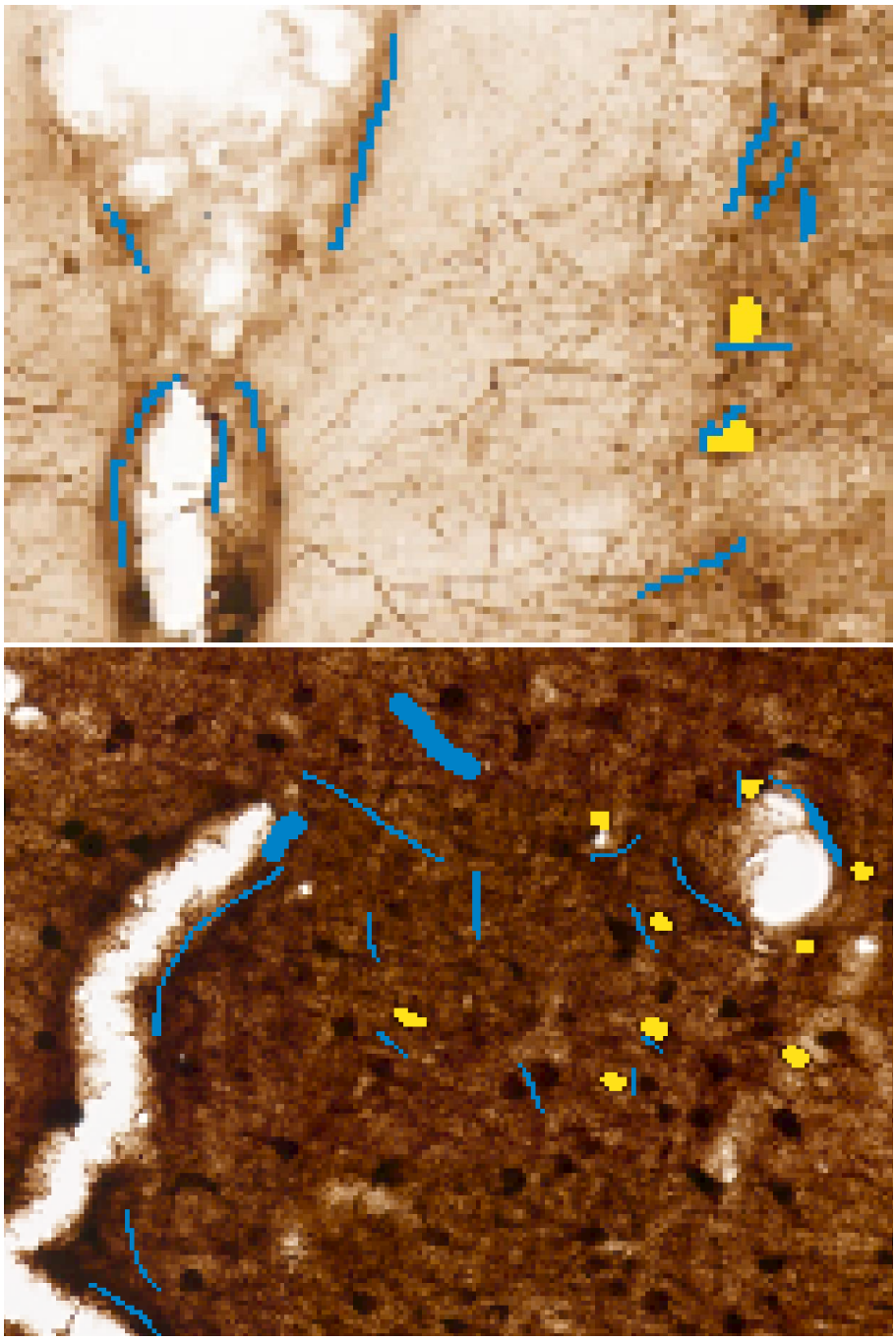
1) Select all features
2) Create two labels: cells and background
3) Use the thinnest brush stroke (1 px)
4) Label some clear examples of cells, and label directly adjacent pixels as background.



Step 2 illustration

5) Label some cells that are closely placed, and label the area between them as background

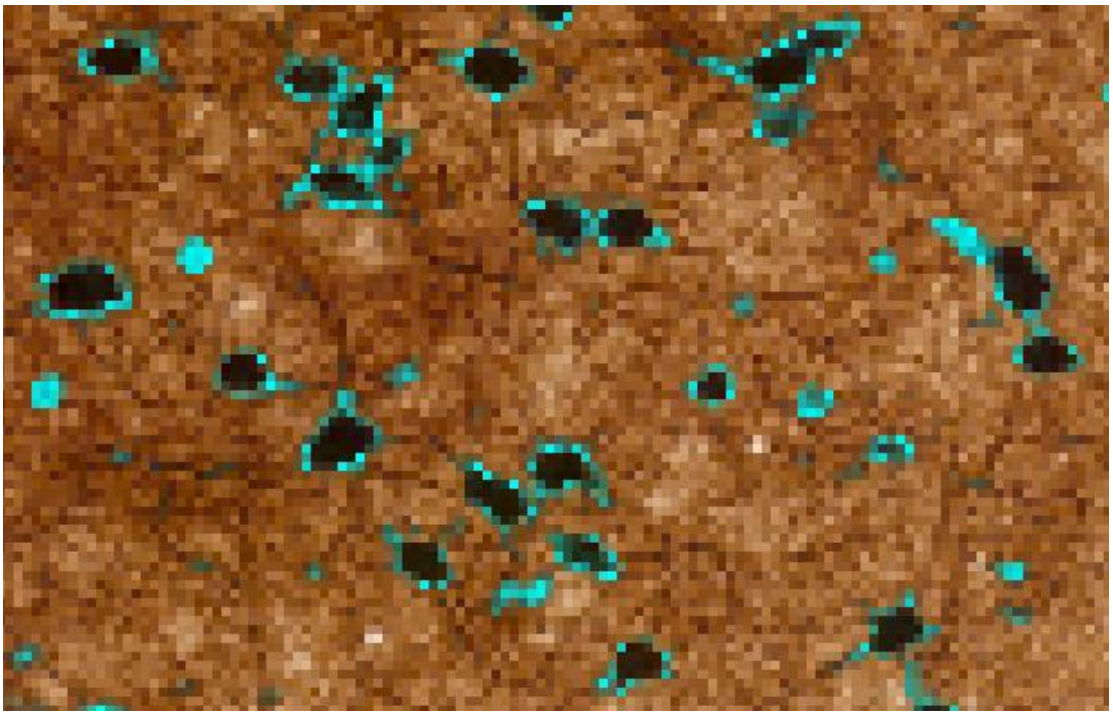
6) Place some background labels in appropriate areas, especially in regions with a lot of stained neuropil and around ventricles where staining is strong but not representing cells.



Step 3-4 illustration

7) Make sure to do steps 3-5 for areas of different intensities and across anatomical regions (cortex, hippocampus, striatum) and training sections. Use a bigger (5 or 7 px) brush to label more extensive areas of background.

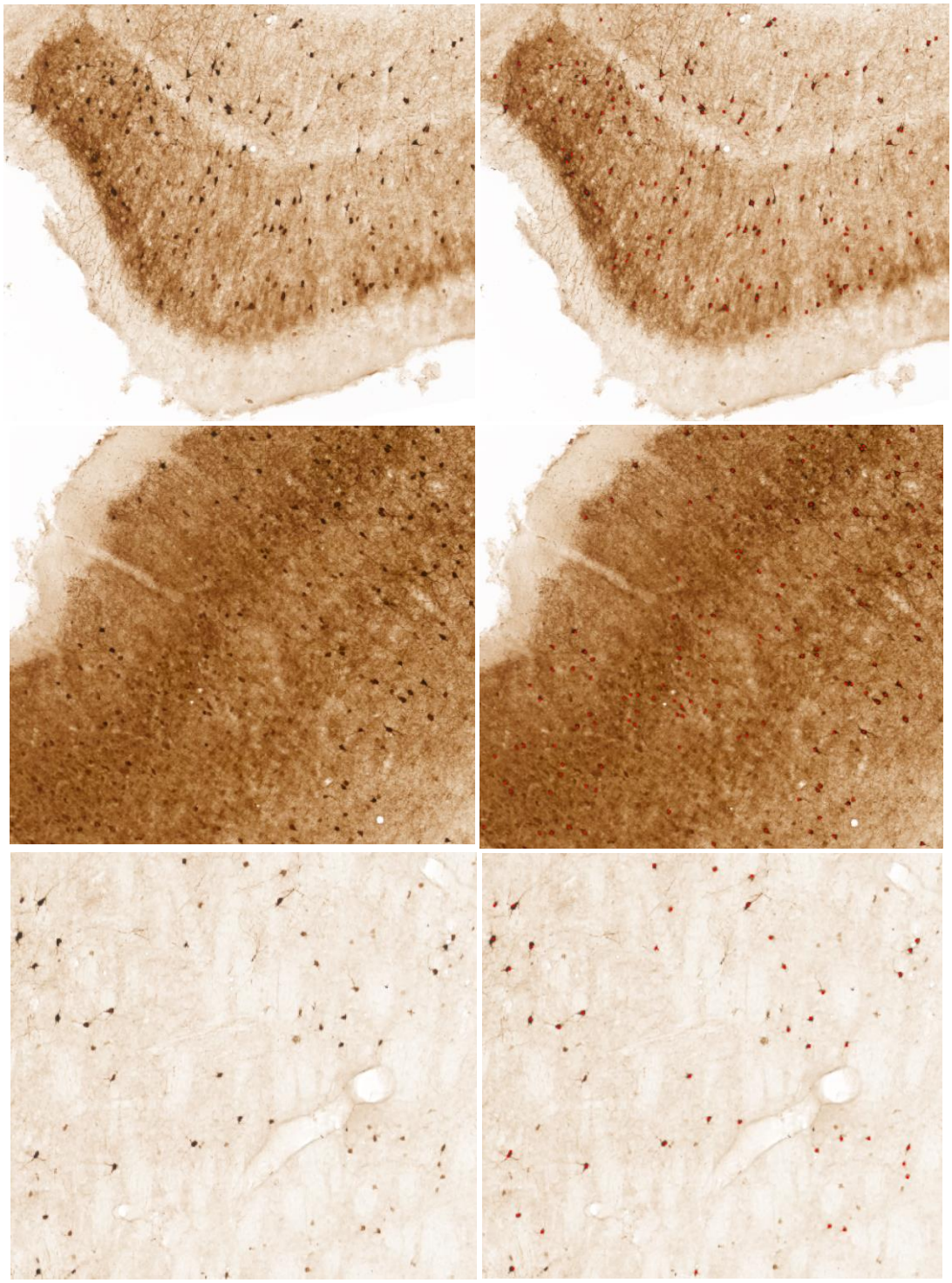
8) Use live update and toggle the uncertainty filter to identify objects for which the classifier needs further input
9) Label cells for which the classifier is uncertain until the uncertainty is restricted to haloes surrounding cells, i.e. until there is little uncertainty in the center of the objects. Make sure to label areas of background at the same time.



Uncertainty filter applied. Cells are recognized with confidence, despite uncertainty at their edges.

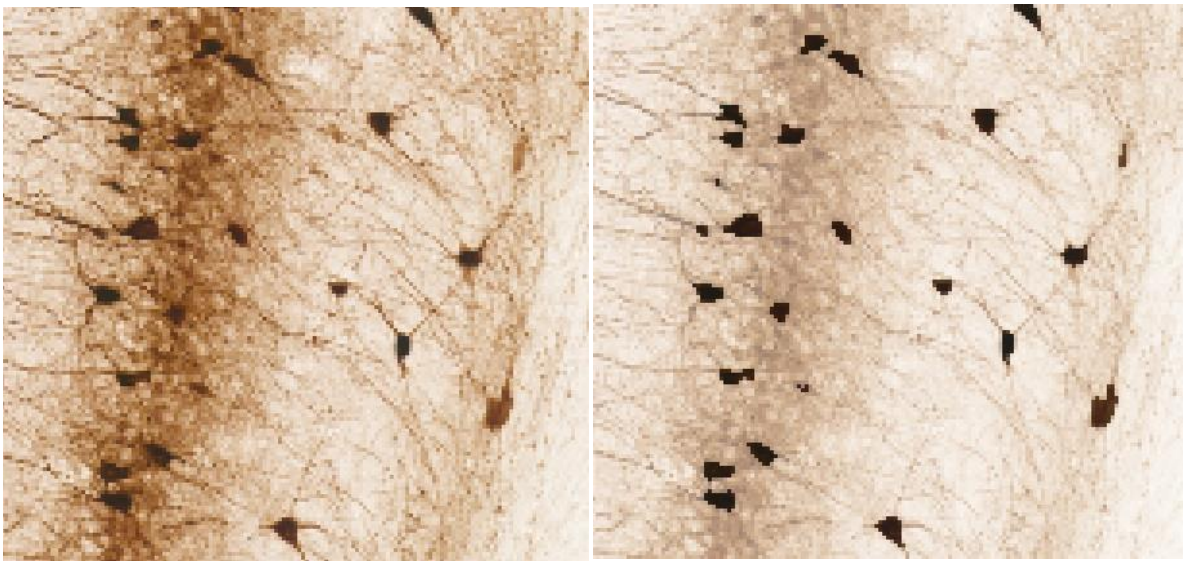
10) Repeat step 8 across all training images, and make sure to visit representative areas (including cortex, striatum, hippocampus, globus pallidus, basal forebrain, olfactory bulb).
11) **See target segmentation images in Navigator to evaluate whether results are satisfactory.** Go through all training images with reference to the segmented training images. If you have a lot of undesired pixels segmented as cells, label more background pixels. Note that a pixel size filter of 4 pixels will be applied to the segmentations, so fragments smaller than this will not affect the results.

Examples of manually annotated cells

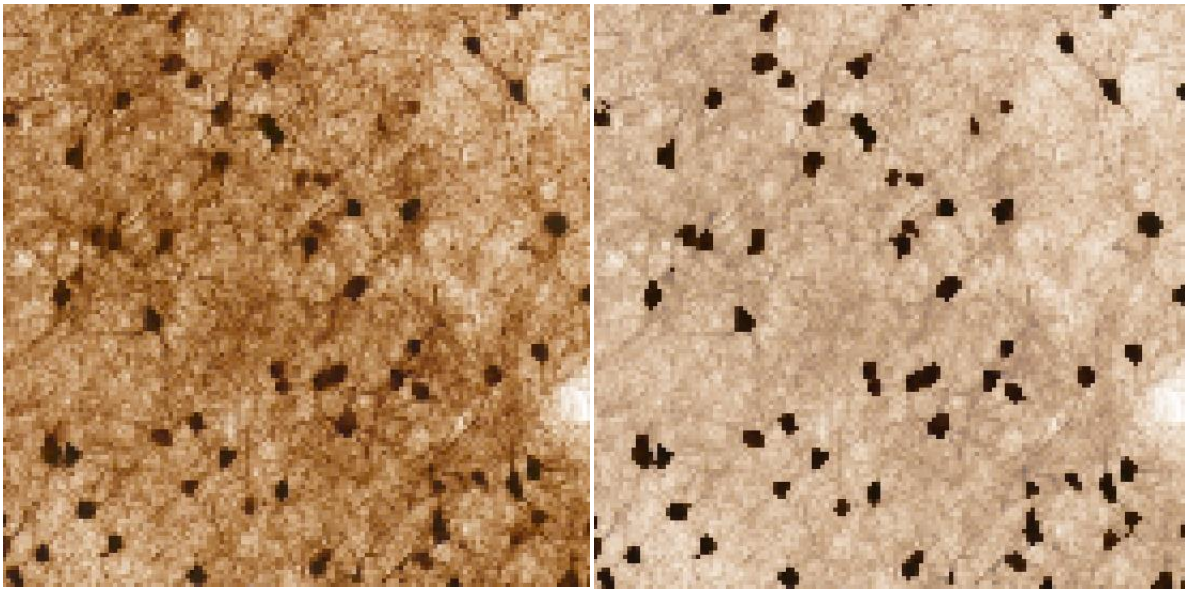


Examples images (left) and the same images with manually annotated cells. The examples are from medial entorhinal cortex (upper), neocortex (middle) and striatum (lower) from s028 of the training images. Very lightly stained objects, or objects that do not have a clear cell shaped outline easily distinguished from background staining, are here not considered to represent cell profiles.

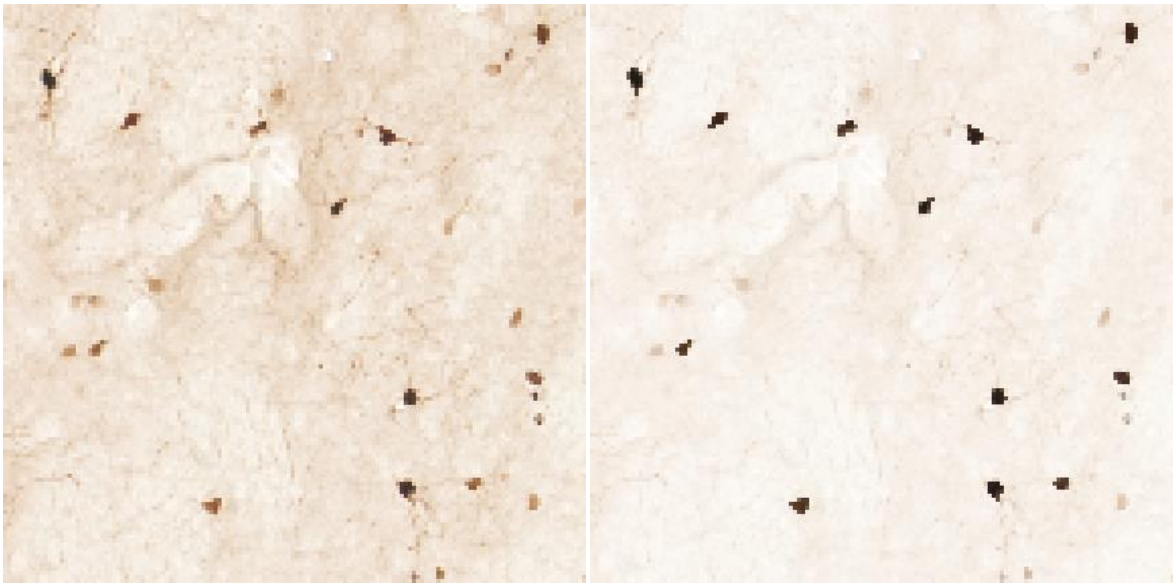
Examples of target segmentation results



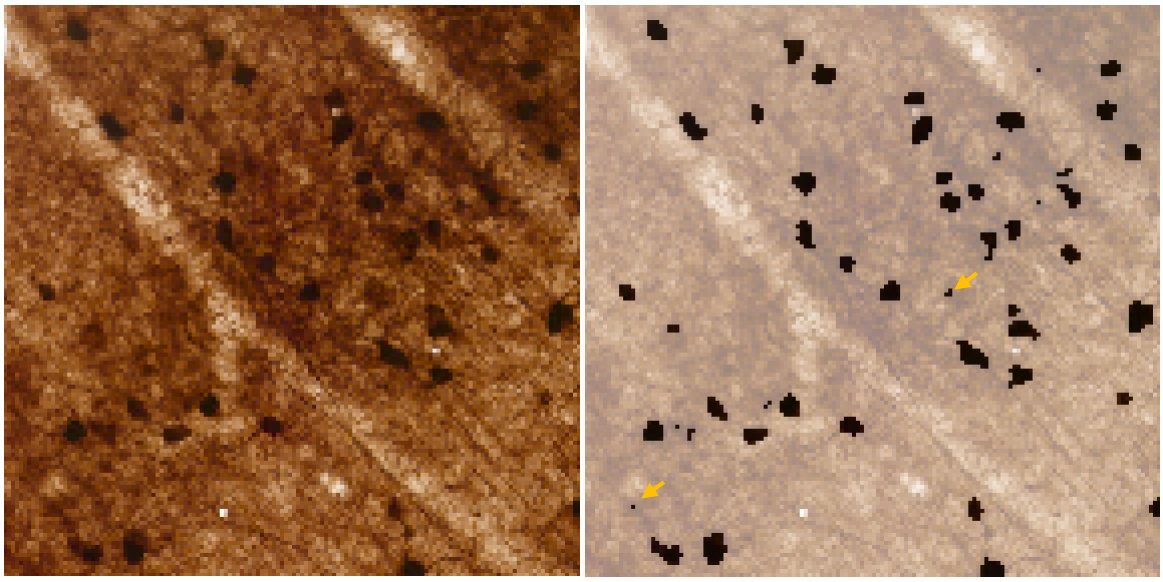
Hippocampus, s014



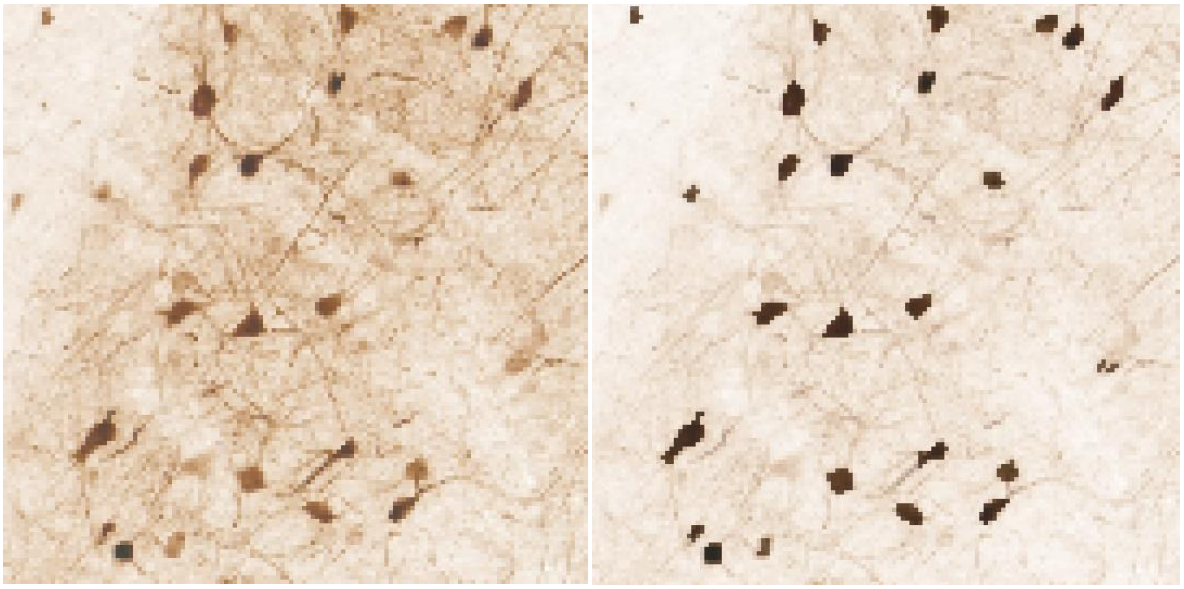
Cortex, s014



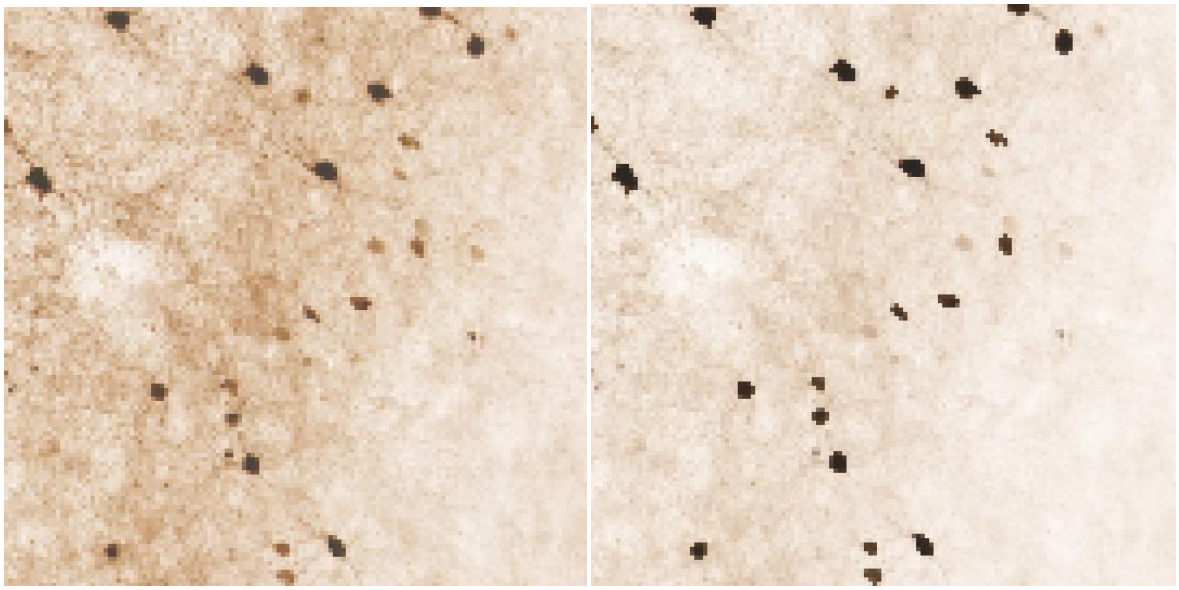
Caudoputamen, s025



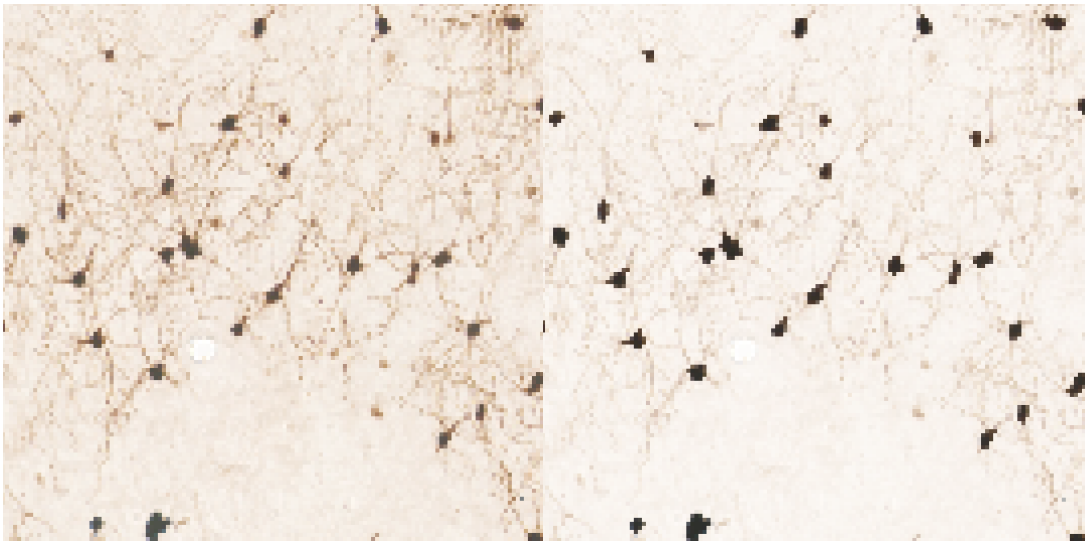
Cortex, s025. Yellow arrows point to examples of fragments that will be filtered out later based on size and thus will not affect count results.



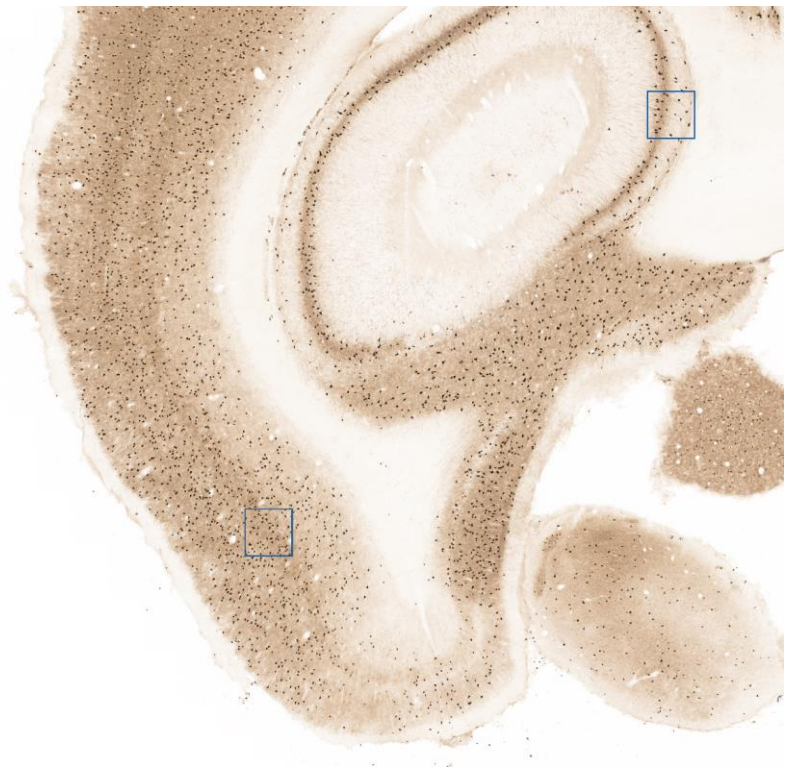
Globus pallidus, s038



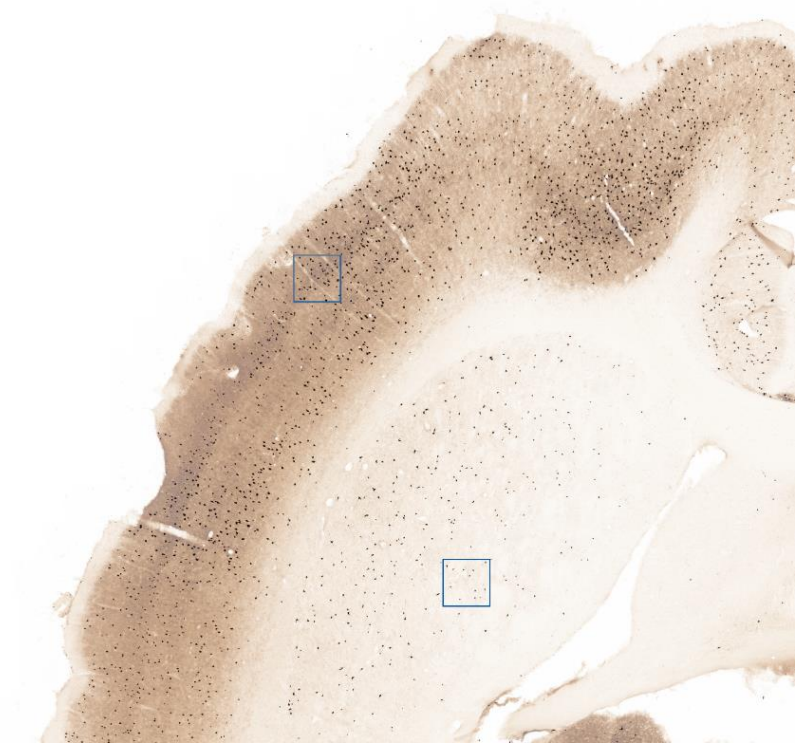
Cortex, s038



Olfactory bulb, s038



Hippocampus and cortex parts shown in examples above, s014



Caudoputamen and cortex parts shown in examples above, s025

Globus pallidus, cortex, and olfactory bulb parts shown in examples above, s038

Supplementary material

Bjerke, IE; Yates, S; Laja, A; Witter, MP; Puchades, MA; Bjaalie, JG; Leergaard, TB*
*Correspondence: Dr. Trygve B. Leergaard: t.b.leergaard@medisin.uio.no

This document contains the data related to our inter- and intrarater reliability testing (see publication for details).

Sheet name	Description
Interrater reliability	*This sheet contains the object counts obtained by quantifying segmentations made by six different researchers (R1-R6). Counts are listed per region in the Waxholm Space atlas of the Sprague-Dawley rat brain (using version 2 of the delineations).*
Intrarater reliability	*This sheet contains the object counts obtained by quantifying segmentations made by one researcher at three different time points (T1-T3). Counts are listed per region in the Waxholm Space atlas of the Sprague-Dawley rat brain (using version 2 of the delineations).*
Definitions major brain regions	*This sheet contains the definition of regions from the Waxholm Space atlas of the Sprague-Dawley rat brain (version 2 of the delineations) into larger groups of "major brain regions". These regions were the ones used for summarising and visualising reliability data in this publication.*

Interrater reliability							
		Object count					
Region ID	Region Name	R1	R2	R3	R4	R5	R6
0	Clear Label	22140	28102	22219	17450	10513	11478
1	descending corticofugal pathways	3022	3284	3250	2307	1696	2000
2	substantia nigra	2504	2671	2590	2397	1951	2012
3	subthalamic nucleus	12	15	13	12	12	6
4	molecular layer of the cerebellum	63909	85881	57788	54166	27070	24618
5	granule cell level of the cerebellum	68728	90728	64816	59592	30197	27636
6	alveus of the hippocampus	8	7	10	1	1	2
7	inferior cerebellar peduncle	188	261	197	150	83	89
10	cingulate cortex, area 2	4881	5036	5069	4586	4309	4477
30	striatum	15569	15809	16528	14220	12287	12348
31	globus pallidus	3869	4073	4070	2498	2228	2658
32	entopeduncular nucleus	174	174	179	77	101	125
33	ventricular system	1192	1298	1176	655	407	552
34	medial lemniscus	274	402	343	304	138	117
35	facial nerve	152	163	156	147	107	115
36	anterior commissure, anterior part	127	130	136	129	81	83
37	anterior commissure, posterior part	78	82	86	69	48	55
38	ventral hippocampal commissure	50	62	58	30	22	29
39	thalamus	16592	18857	18255	14858	10643	11125
40	septal region	758	799	845	609	396	441
41	optic nerve	0	0	0	0	0	0
42	optic tract and optic chiasm	128	153	145	65	62	83
43	pineal gland	144	172	127	95	70	75
44	inner ear	75	107	81	83	40	33
45	spinal cord	0	0	0	0	0	0
46	commissure of the superior colliculus	3	5	3	3	1	2
47	brainstem	44932	52637	47519	41627	28127	28458
48	hypothalamic region	1309	1473	1552	1480	749	745
49	inferior colliculus	9318	11504	10251	10310	5699	4458
50	superficial gray layer of the superior colliculus	2841	3091	3408	2291	1807	1807
51	periaqueductal gray	1731	1902	1858	1405	1051	1129
52	fornix	55	56	64	31	21	23
53	mammillothalamic tract	64	71	71	63	45	42
54	commissural stria terminalis	1	1	1	1	1	1
55	deeper layers of the superior colliculus	7944	8987	8606	7501	6027	6282
56	periventricular gray	1846	2563	1942	1967	825	756
57	genu of the facial nerve	113	125	137	131	70	50
58	pontine nuclei	645	741	648	600	343	285
59	fimbria of the hippocampus	251	244	278	110	77	105
60	fasciculus retroflexus	21	27	27	14	8	12
61	stria medullaris of the thalamus	108	114	131	79	66	72
62	stria terminalis	243	272	260	124	103	140
63	posterior commissure	16	11	16	4	4	9

64	glomerular layer of the accessory olfactory bulb	21	28	29	9	10	12
65	glomerular layer of the olfactory bulb	470	512	567	367	266	276
66	olfactory bulb	14670	15540	16108	11149	10808	11288
67	corpus callosum and associated subcortical white matter	4567	4708	4900	3794	3402	3536
68	brachium of the superior colliculus	9	9	12	4	1	4
69	commissure of the inferior colliculus	78	94	91	78	46	43
70	central canal	0	0	0	0	0	0
71	interpeduncular nucleus	184	235	189	176	96	102
72	ascending fibers of the facial nerve	46	44	49	47	28	25
73	anterior commissure	50	53	55	49	32	35
74	inferior olive	1	1	1	0	1	0
75	spinal trigeminal nuclus	2880	3181	3049	2716	2287	2361
76	spinal trigeminal tract	2686	3651	2772	2369	1593	1632
77	frontal associiation cortex	1678	1956	1837	1534	1135	1259
78	middle cerebellar peduncle	1430	1710	1562	1248	923	865
79	transverse fibers of the pons	273	318	290	263	157	141
80	habenular commissure	1	1	1	1	0	0
81	nucleus of the stria medullaris	12	13	15	7	3	3
82	basal forebrain region	9980	10403	10645	7766	6288	6819
83	supraoptic decussation	74	74	76	29	35	45
84	medial lemniscus decussation	0	0	0	0	0	0
85	pyramidal decussation	0	0	0	0	0	0
92	neocortex	283776	300515	298871	273748	246212	243885
93	bed nucleus of the stria terminalis	88	98	105	72	59	66
94	pretectal region	3771	4142	4043	3425	2835	3005
95	cornu ammonis 3	4300	4731	4735	3374	3108	3357
96	dentate gyrus	3392	3619	3698	2294	2173	2399
97	cornu ammonis 2	888	1021	1035	850	714	748
98	cornu ammonis 1	7394	7891	8382	6751	6157	6349
99	fasciola cinereum	807	885	908	714	645	664
100	subiculum	5377	5658	5659	4864	4594	4757
108	postrhinal cortex	3234	3531	3496	2949	2528	2571
109	presubiculum	6296	6578	6598	6064	5746	5900
110	parasubiculum	3699	3956	3961	3622	3148	3154
112	perirhinal area 35	644	690	743	511	415	427
113	perirhinal area 36	1916	2323	2154	1779	1270	1352
114	entorhinal cortex	8118	8567	8602	7563	6821	6986
115	lateral entorhinal cortex	3134	3254	3391	2619	2194	2204

Summary of object counts per major brain region*

	R1	R2	R3	R4	R5	R6
hippocampus	32 153	34 339	34 976	28 533	26 285	27 328
parahippocampal region	17 046	18 365	18 386	15 421	13 228	13 540
striatum	15 569	15 809	16 528	14 220	12 287	12 348
globus pallidus	4 043	4 247	4 249	2 575	2 329	2 783

hypothalamus	1 309	1 473	1 552	1 480	749	745
substantia nigra	2 504	2 671	2 590	2 397	1 951	2 012
basal forebrain and septal regions	10 738	11 202	11 490	8 375	6 684	7 260
olfactory regions	15 161	16 080	16 704	11 525	11 084	11 576
superior colliculus	10 797	12 092	12 029	9 799	7 836	8 095
inferior colliculus	9 396	11 598	10 342	10 388	5 745	4 501
corpus callosum and associated subcortical white matter	4 567	4 708	4 900	3 794	3 402	3 536
other midbrain	3 871	4 253	4 163	3 504	2 897	3 074
brainstem	55 077	64 734	58 238	50 703	34 770	35 150
neocortex	290 335	307 507	305 777	279 868	251 656	249 621

*** Cerebellum and thalamus data are excluded because the classifiers were not trained in these areas due to oversaturated labelling**

Intrarater reliability				
		Object count		
Region ID	Region Name	T1	T2	T3
0	Clear Label	22140	25354	22179
1	descending corticofugal pathways	3022	3102	2908
2	substantia nigra	2504	2144	2293
3	subthalamic nucleus	12	10	9
4	molecular layer of the cerebellum	63909	79823	55186
5	granule cell level of the cerebellum	68728	83789	58308
6	alveus of the hippocampus	8	6	10
7	inferior cerebellar peduncle	188	211	166
10	cingulate cortex, area 2	4881	4938	4954
30	striatum	15569	15093	17058
31	globus pallidus	3869	3558	3835
32	entopeduncular nucleus	174	150	157
33	ventricular system	1192	1205	1150
34	medial lemniscus	274	249	184
35	facial nerve	152	119	137
36	anterior commissure, anterior part	127	120	150
37	anterior commissure, posterior part	78	73	95
38	ventral hippocampal commissure	50	54	58
39	thalamus	16592	17246	15595
40	septal region	758	677	844
41	optic nerve	0	0	0
42	optic tract and optic chiasm	128	146	132
43	pineal gland	144	99	102
44	inner ear	75	89	66
45	spinal cord	0	0	0
46	commissure of the superior colliculus	3	2	3
47	brainstem	44932	39104	37480
48	hypothalamic region	1309	1450	1436
49	inferior colliculus	9318	11069	8733
50	superficial gray layer of the superior colliculus	2841	2746	2915
51	periaqueductal gray	1731	1595	1614
52	fornix	55	44	55
53	mammillothalamic tract	64	68	65
54	commissural stria terminalis	1	1	1
55	deeper layers of the superior colliculus	7944	6942	7593
56	periventricular gray	1846	2541	1447
57	genu of the facial nerve	113	172	109
58	pontine nuclei	645	530	467
59	fimbria of the hippocampus	251	250	261
60	fasciculus retroflexus	21	19	21
61	stria medullaris of the thalamus	108	122	125
62	stria terminalis	243	257	226
63	posterior commissure	16	21	17
64	glomerular layer of the accessory olfactory bulb	21	29	31

65	glomerular layer of the olfactory bulb	470	514	579
66	olfactory bulb	14670	14699	15818
67	corpus callosum and associated subcortical white matter	4567	4388	4750
68	brachium of the superior colliculus	9	11	11
69	commissure of the inferior colliculus	78	87	74
70	central canal	0	0	0
71	interpeduncular nucleus	184	109	120
72	ascending fibers of the facial nerve	46	40	37
73	anterior commissure	50	49	65
74	inferior olive	1	1	1
75	spinal trigeminal nuclus	2880	2707	2767
76	spinal trigeminal tract	2686	2570	2310
77	frontal associiation cortex	1678	1461	1683
78	middle cerebellar peduncle	1430	1484	1319
79	transverse fibers of the pons	273	229	208
80	habenular commissure	1	1	1
81	nucleus of the stria medullaris	12	8	14
82	basal forebrain region	9980	8652	9549
83	supraoptic decussation	74	75	71
84	medial lemniscus decussation	0	0	0
85	pyramidal decussation	0	0	0
92	neocortex	283776	277664	284307
93	bed nucleus of the stria terminalis	88	90	102
94	pretectal region	3771	3357	3583
95	cornu ammonis 3	4300	4069	4273
96	dentate gyrus	3392	3080	3268
97	cornu ammonis 2	888	909	916
98	cornu ammonis 1	7394	7261	7498
99	fasciola cinereum	807	805	789
100	subiculum	5377	5198	5351
108	postrhinal cortex	3234	3071	3317
109	presubiculum	6296	6361	6432
110	parasubiculum	3699	3809	3817
112	perirhinal area 35	644	601	720
113	perirhinal area 36	1916	1668	1977
114	entorhinal cortex	8118	7956	8279
115	lateral entorhinal cortex	3134	2925	3305

Summary of object counts per major brain region*

hippocampus	32 153	31 492	32 344
parahippocampal region	17 046	16 221	17 598
striatum	15 569	15 093	17 058
globus pallidus	4 043	3 708	3 992
hypothalamus	1 309	1 450	1 436
substantia nigra	2 504	2 144	2 293
basal forebrain and septal regions	10 738	9 329	10 393

olfactory regions	15 161	15 242	16 428
superior colliculus	10 797	9 701	10 522
inferior colliculus	9 396	11 156	8 807
corpus callosum and associated subcortical white matter	4 567	4 388	4 750
other midbrain	3 871	3 455	3 699
brainstem	55 077	48 762	46 615
neocortex	290 335	284 063	290 944

*** Cerebellum and thalamus data are excluded because the classifiers were not trained in these areas due to oversaturated labelling**

Definitions of major brain regions		
Major brain region	**Subregions included**	**Subregion IDs**
hippocampus	cornu ammonis 3; dentate gyrus; cornu ammonis 2; cornu ammonis 1; fasciola cinereum; subiculum; presubiculum; parasubiculum	95; 96; 97; 98; 99; 100; 109; 110
parahippocampal region	postrhinal cortex; perirhinal area 35; perirhinal area 36; entorhinal cortex; lateral entorhinal cortex	108; 112; 113; 114; 115
striatum	striatum	30
globus pallidus	globus pallidus; entopeduncular nucleus	31; 32
hypothalamus	hypothalamic region	48
substantia nigra	substantia nigra	2
basal forebrain and septal regions	septal region; basal forebrain region	40; 82
olfactory regions	olfactory bulb; glomerular layer of the olfactory bulb; glomerular layer of the accessory olfactory bulb	66; 65; 64
superior colliculus	superficial gray layer of the superior colliculus; commissure of the superior colliculus; deeper layers of the superior colliculus; brachium of the superior colliculus	50; 46; 55; 68
inferior colliculus	inferior colliculus; commissure of the inferior colliculus	49; 69
corpus callosum and associated subcortical white matter	corpus callosum and associated subcortical white matter	67
other midbrain	bed nucleus of the stria terminalis; pretectal region; nucleus of the stria medullaris	93; 94; 81
brainstem	inferior cerebellar peduncle; facial nerve; brainstem; periaqueductal gray; genu of the facial nerve; pontine nuclei; ascending fibers of the facial nerve; inferior olive; spinal trigeminal nuclus; spinal trigeminal tract; middle cerebellar peduncle; transverse fibers of the pons	7; 35; 47; 51; 57; 58; 72; 74; 75; 76; 78; 79
neocortex	neocortex; cingulate cortex, area 2; frontal associiation cortex	92; 10; 77

Detailed list of literature sources

Mouse parvalbumin neuron densities							
Source	**Strain**	**Brain region**	**# of animals**	**Counting method**	**Section thickness**	**Density (mm^3)**	**Antibody**
Grunewald_2017 [1]	129/Sv	Amygdala (whole region)	14	Direct count	30	2 649	RRID: AB_2631173
Fasulo_2017 [2]	Other	Basolateral amygdala	4	Stereology	40	1 400	No info
Jinno_2006 [3]	C57BL/6J	CA1	4	Stereology	50	1 405	Kägi et al. (1987)
Neddens_2009 [4]	C57BL/6	CA1	4	Direct count	50	1 158	rabbit polyclonal anti pv, Swant
Pitts_2013 [5]	C57BL/6J	CA1 of dorsal hippocampus	6	Stereology	40	1 480	RRID:AB_10000344
Pitts_2013 [5]	C57BL/6J	CA2/3 of dorsal hippocampus	6	Stereology	40	2 171	RRID:AB_10000344
Jinno_2006 [3]	C57BL/6J	CA3	4	Stereology	50	1 535	Kägi et al. (1987)
Neddens_2009 [4]	C57BL/6	CA3	4	Direct count	50	1 127	rabbit polyclonal anti pv, Swant
Grunewald_2017	129/Sv	CA3	14	Stereology	30	5 796	RRID: AB_2631173
Filice_2016 [6]*	C57BL/6J	Caudoputamen	5	Stereology		897	RRID:AB_10000344
Lauber_2018 [7]*	C57BL/6J	Caudoputamen	6	Stereology		819	RRID:AB_10000344
Lauber_2016 [8]*	C57BL/6J	Caudoputamen	5	Stereology		817	RRID:AB_10000344
Andsberg_2001 [9]*	C57BL/6J	Caudoputamen	7	Stereology	40	864	polyclonal rabbit anti pv
Smith_2008 [10]	Other	Caudoputamen	3	Stereology	50	3 522	No info
Yalcin-Cakmakli_2018 [11]*	C57BL/6J x DBA/2J	Caudoputamen	3	Stereology	30	953	RRID:AB_2631173
Song_2013 [12]*	C57BL/6J	Caudoputamen	12	Stereology	40	612	No info
Filice_2016 [6]*	C57BL/6J	Caudoputamen	5	Stereology		714	RRID:AB_10000344
Ransome_2005 [13]	C57BL/6	Caudoputamen	3	Stereology	30	1 060	mouse anti-parvalbumin, Chemicon
Förster_2008 [14]	C57BL/6J	Caudoputamen	6	Stereology	25	1 430	PARV-19, Sigma
Fasulo_2017 [2]	Other	Dentate gyrus	4	Stereology	40	400	No info
Fasulo_2017 [2]	Other	Dentate gyrus	4	Stereology	40	1 200	No info
Fasulo_2017 [2]	Other	Dentate gyrus	4	Stereology	40	1 100	No info
Jinno_2006 [3]	C57BL/6J	Dentate gyrus	4	Stereology	50	650	Kägi et al. (1987)
Neddens_2009 [4]	C57BL/6	Dentate gyrus	4	Direct count	50	745	rabbit polyclonal anti pv, Swant
Pitts_2013 [5]	C57BL/6J	Dentate gyrus of dorsal hippocampus	6	Stereology	40	493	RRID:AB_10000344
Smith_2008 [10]*	Other	Globus pallidus	3	Stereology	50	40 194	No info
Pitts_2013 [5]	C57BL/6J	Inferior colliculus	6	Stereology	40	4 836	RRID:AB_10000344
Moreno-Gonzalez_2009 [15]	C57BL/6	Lateral entorhinal area	5	Stereology	40	4 300	rabbit polyclonal anti pv, Swant
Parrish-Aungst_2007 [16]	C57BL/6J	Main olfactory bulb	4	Stereology	25	2 324	RRID:AB_10000343
Pirone_2018 [17]	Other	Medial prefrontal cortex (infralimbic)	3	Direct count	20	2 052	RRID: AB_2631173
Pirone_2018 [17]	Other	Medial prefrontal cortex (prelimbic)	3	Direct count	20	3 167	RRID: AB_2631173
Pitts_2013 [5]	C57BL/6J	Medial septum	6	Stereology	40	1 678	RRID:AB_10000344
Smith_2008 [10]*	Other	Nucleus accumbens	3	Stereology	50	2 557	No info
Sanchez-Mejias_2020 [18]	C57BL/6J	Perirhinal area 35	3	Stereology	40	10 765	No info
Sanchez-Mejias_2020 [18]	C57BL/6J	Perirhinal area 36	3	Stereology	40	10 974	No info
Ransome_2005 [13]	C57BL/6	Somatosensory cortex	3	Stereology	30	4 580	mouse anti-parvalbumin, Chemicon
Pitts_2013 [5]	C57BL/6J	Somatosensory cortex	6	Stereology	40	5 329	RRID:AB_10000344
Neddens_2009 [4]	C57BL/6	Subiculum	4	Direct count	50	2 048	rabbit polyclonal anti pv, Swant
Trujillo-Estrada_2014 [19]	C57BL/6	Subiculum	5	Stereology	40	10 721	rabbit polyclonal anti pv, Swant

* These densities were obtained by dividing total number estimates reported in sources by the volume of the region (as defined by the Allen Mouse brain Common Coordinate Framework, version 3, 2017 edition; see main text of paper for details).

Mouse calbindin neuron densities							
Source	**Strain**	**Brain region**	**# of animals**	**Counting method**	**Section thickness**	**Density per mm^3**	**Antibody**
Grunewald_2017 [1]	129/Sv	Amygdala (whole region)	12	Direct count	30	5 005	RRID: AB_10000340
Grunewald_2017 [1]	129/Sv	CA3	14	Stereology	30	945	RRID: AB_10000340
Jinno_2006 [3]	C57BL/6J	CA1	4	Stereology	50	1 555	Pinol et al. (1990)
Jinno_2006 [3]	C57BL/6J	CA3	4	Stereology	50	1 590	Pinol et al. (1990)
Jinno_2006 [3]	C57BL/6J	DG	4	Stereology	50	210	Pinol et al. (1990)
Parrish-Aungst_2007 [16]	C57BL/6J	Main olfactory bulb	4	Stereology	25	10 916	RRID:AB_2721225

Summary of calbindin literature:

Grunewald estimates 5 005 calbindin cells per mm^3 in the amygdala. The average of our corresponding amygdalar regions (lateral, basolateral, basomedial and posterior amygdalar nucleus) is 5 285.

Jinno estimates 1 590 calbindin cells per mm^3 in the CA3 while Grunewald estimates 945 per mm^3. Our estimate is 404 cells per mm^3.

Jinno estimates 1 555 calbindin cells per mm^3 in the CA1. Our estimate is 793.

Parrish-Aungst estimates 10 916 calbindin cells per mm^3 in the main olfactory bulb. Our estimate is 3 517. However, note that they sample only from a subset of the main olfactory bulb, stating that "Sections were randomly selected from the middle third of the olfactory bulb and analyzed only on the medial wall. This strategy was employed to prevent any regional variations or bilateral asymmetry in the bulb".

Rat parvalbumin neuron densities							
Source	**Strain**	**Brain region**	**# of animals**	**Counting method**	**Section thickness**	**Density (mm^3)**	**Antibody**
Wang_2008 [20]	Sprague-Dawley	CA1		Direct count	50	1 600	RRID:AB_477329
Aika_1994 [21]	Wistar	CA1, dorsal	5	Stereology	0,5	1 100	Polyclonal rabbit anti PV antibody
Wang_2008 [20]	Sprague-Dawley	Caudoputamen		Direct count	50	846	RRID:AB_477329
Kaalund_2013 [22]	Lister Hooded	Dentate gyrus		Stereology	60	2 128	No info
Megahed_2015 [23]	Sprague-Dawley	Dentate gyrus	6	Stereology	30	2 125	No info
Shiraki_2016 [24]	Sprague-Dawley	Dentate gyrus hilus	10	Direct count	3	1 461	PARV-19 Millipore
Kaalund_2013 [22]	Lister Hooded	Hippocampus CA		Stereology	60	2 994	No info
Megahed_2015 [23]	Sprague-Dawley	Hippocampus CA	6	Stereology	30	3 064	No info
Barinka_2012 [25]	Wistar	Perirhinal area 35	6	Stereology		3 805	Monoclonal mouse anti PV
Barinka_2012 [25]	Wistar	Perirhinal area 36	6	Stereology		4 336	Monoclonal mouse anti PV

Summary of parvalbumin rat data:

Wang et al. estimated 1 600 parvalbumin neurons per mm^3 in the CA1 of the hippocampus of Sprague-Dawley rats, while Aika estimated 1 100 parvalbumin neurons per mm^3 in the dorsal CA1. Kaalund estimated 2 994 while Megahed estimated 3 064 parvalbumin neurons per mm^3 in Ammon's horn. Our estimate is 866 parvalbumin neurons per mm^3 in the CA1.

Estimates from the dentate gyrus in the literature range from 1 461 to 2 128 parvalbumin neurons per mm^3 (some estimates from hilus region only). Our estimate is 347 parvalbumin neurons per mm^3.

Barinka estimated 3 805 and 4 336 parvalbumin neurons per mm^3 in perirhinal areas 35 and 36, respectively. Our estimates are 1 188 and 1 258. It should be noted that Barinka and colleagues excluded layer I, which contains few to no cells, which may contribute to their relatively higher estimate.

Additional note: In an interesting study, Bezaire & Soltesz [26] used a combination of literature derived data and calculations to arrive at an estimate of parvalbumin neurons in the rat CA1. Their estimate is based on a total number estimate for all neurons in CA1, then the proportion of these that are assumed to be interneurons gave an estimate of total interneuron number. Then this number and the proportion of interneurons thought to express parvalbumin gave an estimate of 10 010 parvalbumin positive neurons unilaterally. Our estimate is 13 798 parvalbumin positive neurons unilaterally in rat CA1.

References:

1. Grünewald B, Lange MD, Werner C, O'Leary A, Weishaupt A, Popp S, et al. Defective synaptic transmission causes disease signs in a mouse model of juvenile neuronal ceroid lipofuscinosis. Elife. 2017;6. doi:10.7554/eLife.28685

2. Fasulo L, Brandi R, Arisi I, La Regina F, Berretta N, Capsoni S, et al. ProNGF drives localized and cell selective parvalbumin interneuron and perineuronal net depletion in the dentate gyrus of transgenic mice. Front Mol Neurosci. 2017;10: 20. doi:10.3389/fnmol.2017.00020

3. Jinno S, Kosaka T. Cellular architecture of the mouse hippocampus: A quantitative aspect of chemically defined GABAergic neurons with stereology. Neuroscience Research. Elsevier; 2006. pp. 229–245. doi:10.1016/j.neures.2006.07.007

4. Neddens J, Buonanno A. Selective populations of hippocampal interneurons express ErbB4 and their number and distribution is altered in ErbB4 knockout mice. Hippocampus. 2009;20: NA-NA. doi:10.1002/hipo.20675

5. Pitts MW, Reeves MA, Hashimoto AC, Ogawa A, Kremer P, Seale LA, et al. Deletion of selenoprotein M leads to obesity without cognitive deficits. J Biol Chem. 2013;288: 26121–26134. doi:10.1074/jbc.M113.471235

6. Filice F, Vörckel KJ, Sungur AÖ, Wöhr M, Schwaller B. Reduction in parvalbumin expression not loss of the parvalbumin-expressing GABA interneuron subpopulation in genetic parvalbumin and shank mouse models of autism. Mol Brain. 2016; 1–17. doi:10.1186/s13041-016-0192-8

7. Lauber E, Filice F, Schwaller B. Dysregulation of Parvalbumin Expression in the Cntnap2 − / − Mouse Model of Autism Spectrum Disorder. Front Mol Neurosci. 2018;11: 1–15. doi:10.3389/fnmol.2018.00262

8. Lauber E, Filice F, Schwaller B. Prenatal Valproate Exposure Differentially Affects Parvalbumin-Expressing Neurons and Related Circuits in the Cortex and Striatum of Mice. Front Mol Neurosci. 2016;9: 1–16. doi:10.3389/fnmol.2016.00150

9. Andsberg G, Kokaia Z, Lindvall O. Upregulation of p75 neurotrophin receptor after stroke in mice does not contribute to differential vulnerability of striatal neurons. Exp Neurol. 2001;169: 351–363. doi:10.1006/exnr.2001.7646

10. Smith KM, Fagel DM, Stevens HE, Maragnoli ME, Rabenstein RL, Ohkubo Y, et al. Deficiency in Inhibitory Cortical Interneurons Associates with Hyperactivity in Fibroblast Growth Factor Receptor 1 Mutant Mice. Biol Psychiatry. 2008;63: 953–962. doi:10.1016/j.biopsych.2007.09.020

11. Yalcin-Cakmakli G, Rose SJ, Villalba RM, Williams L, Jinnah HA, Hess EJ, et al. Striatal Cholinergic Interneurons in a Knock-in Mouse Model of L -DOPA-Responsive Dystonia. Front Syst Neurosci. 2018;12: 1–12. doi:10.3389/fnsys.2018.00028

12. Song C-H, Bernhard D, Bolarinwa C, Hess EJ, Smith Y, Jinnah HA. Subtle microstructural changes of the striatum in a DYT1 knock-in mouse model of dystonia. Neurobiol Dis. 2013;54: 362–71. doi:10.1016/j.nbd.2013.01.008

13. Ransome MI, Turnley AM. Analysis of neuronal subpopulations in mice over-expressing suppressor of cytokine signaling-2. Neuroscience. 2005;132: 673–687. doi:10.1016/j.neuroscience.2004.12.041

14. Förster J. Quantitative morphological analysis of the neostriatum and the cerebellum of tenascin-C deficient mice. University of Hamburg. 2008. Available: https://ediss.sub.uni-hamburg.de/volltexte/2008/3923/pdf/dissertationj.foerster.pdf

15. Moreno-Gonzalez I, Baglietto-Vargas D, Sanchez-Varo R, Jimenez S, Trujillo-Estrada L, Sanchez-Mejias E, et al. Extracellular amyloid-β and cytotoxic glial activation induce significant entorhinal neuron loss in young

PS1M146L/APP751SL mice. J Alzheimer's Dis. 2009;18: 755–776. doi:10.3233/JAD-2009-1192

16. Parrish-Aungst S, Shipley MT, Erdelyi F, Szabo G, Puche AC. Quantitative analysis of neuronal diversity in the mouse olfactory bulb. J Comp Neurol. 2007;501: 825–836. doi:10.1002/cne.21205

17. Pirone A, Alexander JM, Koenig JB, Cook-Snyder DR, Palnati M, Wickham RJ, et al. Social Stimulus Causes Aberrant Activation of the Medial Prefrontal Cortex in a Mouse Model With Autism-Like Behaviors. Front Synaptic Neurosci. 2018;10: 35. doi:10.3389/fnsyn.2018.00035

18. Sanchez-Mejias E, Nuñez-Diaz C, Sanchez-Varo R, Gomez-Arboledas A, Garcia-Leon JA, Fernandez-Valenzuela JJ, et al. Distinct disease-sensitive GABAergic neurons in the perirhinal cortex of Alzheimer's mice and patients. Brain Pathol. 2020;30: 345–363. doi:10.1111/bpa.12785

19. Trujillo-Estrada L, Dávila JC, Sánchez-Mejias E, Sánchez-Varo R, Gomez-Arboledas A, Vizuete M, et al. Early neuronal loss and axonal/presynaptic damage is associated with accelerated amyloid-β accumulation in AβPP/PS1 Alzheimer's disease mice subiculum. J Alzheimer's Dis. 2014;42: 521–541. doi:10.3233/JAD-140495

20. Wang CZ, Yang SF, Xia Y, Johnson KM. Postnatal phencyclidine administration selectively reduces adult cortical parvalbumin-containing interneurons. Neuropsychopharmacology. 2008;33: 2442–2455. doi:10.1038/sj.npp.1301647

21. Aika Y, Ren JQ, Kosaka K, Kosaka T. Quantitative analysis of GABA-like-immunoreactive and parvalbumin-containing neurons in the CA1 region of the rat hippocampus using a stereological method, the disector. Exp Brain Res. 1994;99: 267–276. doi:10.1007/BF00239593

22. Kaalund SS, Riise J, Broberg B V., Fabricius K, Karlsen AS, Secher T, et al. Differential expression of parvalbumin in neonatal phencyclidine-treated rats and socially isolated rats. J Neurochem. 2013;124: 548–557. doi:10.1111/jnc.12061

23. Megahed T, Hattiangady B, Shuai B, Shetty AK. Parvalbumin and neuropeptide Y expressing hippocampal GABA-ergic inhibitory interneuron numbers decline in a model of Gulf War illness. Front Cell Neurosci. 2015;8: 1–12. doi:10.3389/fncel.2014.00447

24. Shiraki A, Tanaka T, Watanabe Y, Saito F, Akahori Y, Imatanaka N, et al. Immunohistochemistry of aberrant neuronal development induced by 6-propyl-2-thiouracil in rats. Toxicol Lett. 2016;261: 59–71. doi:10.1016/j.toxlet.2016.08.019

25. Barinka F, Salaj M, Rybář J, Krajčovičová E, Kubová H, Druga R. Calretinin, parvalbumin and calbindin immunoreactive interneurons in perirhinal cortex and temporal area Te3V of the rat brain: Qualitative and quantitative analyses. Brain Res. 2012;1436: 68–80. doi:10.1016/j.brainres.2011.12.014

26. Bezaire MJ, Soltesz I. Quantitative assessment of CA1 local circuits: knowledge base for interneuron-pyramidal cell connectivity. Hippocampus. 2013;23: 751–785. doi:10.1002/hipo.22141

Search string used for literature search

To search the literature for data to be compared to our estimates, we performed several searches using Google Scholar and PubMed. For each search, all results were screened manually and included if they presented relevant data (total number or density estimate of parvalbumin or calbindin cells in rats or mice).

The following searches were performed:

1. Google Scholar search for RRID:AB_2631173
Performed 14.04.2020
31 results, 2 included

2. Google Scholar search for RRID:AB_477329
Performed 20.04.2020
81 results, none included

3. Google Scholar search for RRID:AB_10000347
Performed 21.04.2020
87 results, none included

4. Google Scholar search for ((rat) OR (mouse)) AND (Swant PV27) AND (stereology OR stereological)
Performed 14.04.2020
18 results, 1 included

5. Google Scholar search for ((rat) OR (mouse)) AND (PARV-19) AND (stereology OR stereological)
Performed 20.04.2020
61 results, 2 included

6. Google Scholar search for ((rat) OR (mouse)) AND (“Swant 300”) AND (stereology OR stereological)
Performed 21.04.2020
21 results, none included

7. PubMed (via Ovid Medline) search for: (rat or mouse or rodent).tw,kf AND stereolog*.tw,kf. AND parvalbumin.tw,kf.
Performed 14.04.2020
59 results, 14 included

8. PubMed (via Ovid Medline) search for: (rat or mouse or rodent).tw,kf AND stereolog*.tw,kf. AND calbindin*.tw,kf.
Performed 21.04.2020
38 results, 3 included