FAKULTÄT FÜR !NFORMATIK
Faculty of Informatics

From Neurons to Behavior: Visual Analytics Methods for Heterogeneous Spatial Big Brain Data

DISSERTATION

zur Erlangung des akademischen Grades

Doktor der Technischen Wissenschaften

eingereicht von

Dipl.-Ing. Florian Johann Ganglberger
Matrikelnummer 0828078

an der Fakultät für Informatik

der Technischen Universität Wien

Betreuung: Ao.Univ.Prof. Dipl.Ing. Dr.techn. Eduard Gröller
Zweitbetreuung: Dipl.-Math.in Dr. Katja Bühler

Diese Dissertation haben begutachtet:

Anna Vilanova	Torsten Wolfgang Kuhlen

Wien, 25. Juni 2019

Florian Johann Ganglberger

FAKULTÄT FÜR !NFORMATIK
Faculty of Informatics

From Neurons to Behavior: Visual Analytics Methods for Heterogeneous Spatial Big Brain Data

DISSERTATION

submitted in partial fulfillment of the requirements for the degree of

Doktor der Technischen Wissenschaften

by

Dipl.-Ing. Florian Johann Ganglberger
Registration Number 0828078

to the Faculty of Informatics

at the TU Wien

Advisor: Ao.Univ.Prof. Dipl.Ing. Dr.techn. Eduard Gröller
Second advisor: Dipl.-Math.in Dr. Katja Bühler

The dissertation has been reviewed by:

Anna Vilanova	Torsten Wolfgang Kuhlen

Vienna, 25th June, 2019

Florian Johann Ganglberger

Erklärung zur Verfassung der Arbeit

Dipl.-Ing. Florian Johann Ganglberger

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Wien, 25. Juni 2019

Florian Johann Ganglberger

Acknowledgements

The publications presented in this thesis as well the write-up of this thesis have been part of my work at the VRVis Research Center in collaboration with the Institute of Molecular Pathology in Vienna. At first I want to thank all people that are part of this great organizations for making this thesis happen. First and foremost, I want to thank Katja Bühler for inviting me to join her research group. She supervised the projects that ultimately led to this thesis. Her involvement in conceiving the methods, writing, and taking care of funding/project acquisition as well as providing guidance and advice cannot be taken for granted. Furthermore, I want to thank Wulf Haubensak. His collaboration represented a cornerstone for many papers that are either included in or were spawned by this thesis. His ambition and commitment was a driving force behind our joint projects. Without his neuroscientific expertise this thesis would not have been possible. Special thanks also to the supervisor of this thesis, Eduard Gröller. He was never short of good advice or an occasional exhilarating joke that greatly supported me during my doctoral studies.

I also want to thank my colleagues and co-authors of papers, especially Joanna Kaczanowska and Nicolas Swoboda. Joanna's endurance and ingenuity as a PhD student was exceptional, and her contribution was definitely an enrichment for our joint projects. Nicolas offered me his advice for web-development whenever I got stuck with coding. His coffee was a delight, keeping me awake during long crunch-times before publication deadlines. I also want to thank Markus Töpfer and Florian Schulze for lending me their technical expertise, as well as many thanks to Sophia Ulonska and Franziska Steyer-Beerman for proofreading.

My thanks also goes to the funding agencies. This work is the result of a joint VRVis/IMP (Research Institute of Molecular Pathology) project supported by Grant 852936 of the Austrian FFG Funding Agency. VRVis is funded by BMVIT, BMWFW, Styria, SFG, and Vienna Business Agency in the scope of COMET - Competence Centers for Excellent Technologies (854174) which is managed by FFG.

Special thanks goes to my parents who laid the foundation for this thesis more than 30 years ago, and as a consequence acted as my primary funding agency prior to this thesis. They also provided emotional support that, to my better knowledge, no Austrian/European funding agency offers. I also want to thank my sister, the weekly

Club-Berlin Pubquizrunde, the Perger-Partie, the Cantina Band, the Rapid-Abo-Runde, and all my other friends for giving me my inner balance to work on this thesis.

Finally, I want to thank my significant other, Isabella Tutzer, for her patience and support. You gave me the strength to make all of this possible.

"You keep working. You keep trying. You keep failing. Until one day (...) the failure ends. That's all success feels like. It's not triumphant. It's not glorious. It's just a relief. You finally stopped failing." - All Our Wrong Todays by Elan Mastai

Kurzfassung

Fortschritte im Bereich der neurobiologischen Bildgebung ermöglichten Gehirnforschungsinitiativen die Erstellung gewaltiger Datenmengen, die dazu verwendet werden können, mentale Prozesse und biologische Vorgänge besser zu verstehen. Zu diesen Prozessen zählen auch neurologische Erkrankungen wie z.B. Alzheimer, Autismus und Angststörungen. Die Erforschung der Zusammenhänge von Genen, Gehirnschaltkreisen und Verhalten stellt hierbei ein Schlüsselelement dar. Dafür ist eine gemeinsame Analyse von heterogenen räumlichen Gehirndaten notwendig, zu denen 3D-Bildgebungsdaten, anatomische Daten und Gehirnnetzwerke verschiedener Größen, Auflösungen sowie Modalitäten zählen. Da aktuelle Bildgebungsplattformen hohe Durchsatzraten erzielen, werden Daten generiert, deren Größe und Komplexität den bisherigen Stand der Technik um mehrere Größenordnungen übertreffen. Gegenwärtige analytische Arbeitsabläufe umfassen eine zeitaufwändige manuelle Datenaggregierung und umfangreiche computerbasierte Analysen in skriptbasierten Toolboxen. Visuell-analytische Methoden zur Erforschung großer Gehirndaten können Neurowissenschaftler dabei unterstützen, sich mehr auf das Verständnis der Daten konzentrieren zu können.

Diese Dissertation beinhaltet mehrere Beiträge, die sich mit diesem Problem befassen. Der erste Beitrag beschäftigt sich mit einer computerbasierten Methode, mit der genetische Informationen mit räumlichen Genexpressionsdaten und Konnektivitätsdaten verbunden werden sollen, um so funktionelle, neuroanatomische Karten zu erstellen. Die entstehenden Karten zeigen Gehirnregionen, die mit spezifischen Gehirnfunktionen oder Verhaltensweisen in Verbindung gebracht werden können. Mit diesem Ansatz konnte eine bisher unbekannte funktionelle Neuroanatomie von Verhaltensweisen, die mehrere Gene betreffen, prognostiziert werden. Dabei wurde festgestellt, dass sich funktionell zusammengehörige Gene nicht zufällig, sondern in spezifischen Gehirnnetzwerken anhäufen. Der zweite Beitrag befasst sich mit einer Datenstruktur, die eine interaktive Erforschung von umfangreichen Gehirnnetzwerken mit Milliarden von Verbindungen ermöglicht. Dabei wird die hierarchische und räumliche Organisation der Daten genutzt, um eingehende/ausgehende Verbindungen von beliebigen Regionen verschiedener Größenordnungen auf Abruf zu ermöglichen, was ansonsten die Grenzen von handelsüblichen Computern überschreiten würde. Diese Datenstruktur wird im dritten Beitrag dazu verwendet, ein neuartiges, webbasiertes Framework zur Erforschung von neurobiologischen Bildgebungs- und Konnektivitätsdaten unterschiedlicher Typen, Modalitäten und Größenordnungen zu entwickeln. Hierbei wird ein abfragebasiertes Interaktionsschema verwendet, um drei-

dimensionale, räumliche Genexpression und verschiedene Arten von Konnektivität zu untersuchen. Dies ermöglicht die interaktive Sezierung von Gehirnnetzwerken in Bezug auf deren genetische Zusammensetzung in Echtzeit. Um Neurowissenschaftlern den Vergleich von multimodalen Netzwerken in unterschiedlichen Maßstäben zu ermöglichen, werden die Daten im Zusammenhang mit der hierarchischen Organisation von gebräuchlichen anatomischen Atlanten dargestellt. Des Weiteren wird das Framework für kollaboratives Arbeiten mithilfe von teilbaren und nachvollziehbaren Arbeitsabläufen im Web optimiert.

Die im Rahmen dieser Dissertation präsentierten Ansätze können Neurowissenschaftler dabei unterstützen, die funktionelle Organisation des Gehirns über einfache anatomische Domänen hinaus zu verfeinern, und das Wissen darüber, wie Gene unseren Verstand beeinflussen, zu erweitern.

Abstract

Advances in neuro-imaging have allowed big brain initiatives and consortia to create vast resources of brain data that can be mined for insights into mental processes and biological principles. Research in this area does not only relate to mind and consciousness, but also to the understanding of many neurological disorders, such as Alzheimer's disease, autism, and anxiety. Exploring the relationships between genes, brain circuitry, and behavior is therefore a key element in research that requires the joint analysis of a heterogeneous set of spatial brain data, including 3D imaging data, anatomical data, and brain networks at varying scales, resolutions, and modalities. Due to high-throughput imaging platforms, this data's size and complexity goes beyond the state-of-the-art by several orders of magnitude. Current analytical workflows involve time-consuming manual data aggregation and extensive computational analysis in script-based toolboxes. Visual analytics methods for exploring big brain data can support neuroscientists in this process, so they can focus on understanding the data rather than handling it.

In this thesis, several contributions that target this problem are presented. The first contribution is a computational method that fuses genetic information with spatial gene expression data and connectivity data to predict functional neuroanatomical maps. These maps indicate, which brain areas might be related to a specific function or behavior. The approach has been applied to predict yet unknown functional neuroanatomy underlying multigeneic behavioral traits identified in genetic association studies and has demonstrated that rather than being randomly distributed throughout the brain, functionally-related gene sets accumulate in specific networks. The second contribution is the creation of a data structure that enables the interactive exploration of big brain network data with billions of edges. By utilizing the resulting hierarchical and spatial organization of the data, this approach allows neuroscientists on-demand queries of incoming/outgoing connections of arbitrary regions of interest on different anatomical scales. These queries would otherwise exceed the limits of current consumer level PCs. The data structure is used in the third contribution, a novel web-based framework to explore neurobiological imaging and connectivity data of different types, modalities, and scale. It employs a query-based interaction scheme to retrieve 3D spatial gene expressions and various types of connectivity to enable an interactive dissection of networks in real-time with respect to their genetic composition. The data is related to a hierarchical organization of common anatomical atlases that enables neuroscientists to compare multimodal networks on

different scales in their anatomical context. Furthermore, the framework is designed to facilitate collaborative work with shareable comprehensive workflows on the web.

As a result, the approaches presented in this thesis may assist neuroscientists to refine their understanding of the functional organization of the brain beyond simple anatomical domains and expand their knowledge about how our genes affect our mind.

Contents

Kurzfassung **ix**

Abstract **xi**

Contents **xiii**

1 Introduction **1**
1.1 Thesis Challenges and Goals 2
1.2 Scope and Contributions 3
1.3 Contributions of Co-Authors 8
1.4 Thesis Structure 9

2 Background and Related Work **11**
2.1 Spatial Brain Data to Explore the Relationship between Genes, Brain, and Behavior 13
2.2 Visual Analytics Tools for Spatial Brain Data 18

3 Detailed Contributions **31**
3.1 Paper A: Predicting Functional Neuroanatomical Maps 31
3.2 Paper B: A Data Structure for Big Brain Networks 37
3.3 Paper C and D: BrainTrawler 41

4 Conclusion **47**
4.1 Impact 48
4.2 Future Work 49

A Predicting functional neuroanatomical maps from fusing brain networks with genetic information **51**

B A data structure for real-time aggregation queries of big brain networks **73**

C Iterative exploration of big brain network data **117**

D BrainTrawler: A visual analytics framework for iterative exploration of heterogeneous big brain data **129**

List of Figures **153**

Bibliography **157**

Curriculum Vitae **169**

CHAPTER 1

Introduction

Behavioral neuroscience is the study of mental processes and their biological principles. Scientific advances in this area do not only contribute to the knowledge about the mind and consciousness [Koc04], but also to the understanding of many medical disorders, such as Alzheimer's disease, schizophrenia, autism, anxiety, etc. [Kin14]. To find neuronal representations of behavioral patterns, a highly heterogeneous set of neurobiological spatial data, including imaging data (3D imaging data on voxel/brain region-level) and networks (structural, functional, and genetic relations between brain voxels/regions) has to be investigated and put into the context of their mental effects. Recent advances in neuroimaging allowed big brain initiatives and consortia to create vast resources [all19, hum19, VESB$^+$13, mPlDI$^+$16] of this data that covers genes, function, anatomy, and behavior.

The entry point of many neuroscientific experimental workflows are so-called candidate regions (i.e., brain regions that are part of a specific neuronal circuit that is responsible for a brain function or behavior). Relating these candidate regions to genes can then be used to study the circuit dynamics targeted by psychoactive drugs [GPB$^+$18a]. Thus, the knowledge of where genes affect the brain is a first step to relating them with a function. These affected brain regions are rather broadly defined, but are well known, for example, as primary gene expression sites [LHA$^+$07], that are sites where the gene creates products such as proteins (Figure 1.1A) or brain regions that are connected to these sites [OHN$^+$14] (Figure 1.1B and C). Hence, the effect of a gene on behavior/function is not only limited by its expressing brain regions, but might also influence brain regions that interact with it in a neuronal circuit or brain network, such as in a ligand-receptor binding [YW04]. These interactions can be of various types. Comparing them can reveal the dynamics of neural circuits. For example, the task-fMRI connectivity (fMRI signal correlation over time during a task) indicates the functional relationship between two brain regions, while the genetic connectivity (correlation of gene expression) could reveal that the activity between these regions underlies completely different molecular mechanisms. Directional

structural connectivity can even show the flow of information for (unidirectional) fMRI connectivity [GSB+15]. Hence, the central research question of this thesis is: *How to explore spatial gene expression and multimodal brain connectivity data in a joint manner - i.e., how certain brain areas are structurally, functionally, or genetically related?*

1.1 Thesis Challenges and Goals

In recent years, spatial neurobiological data has been generated in an unprecedented scale and resolution. This includes vast spatial gene expression resources (the representation of tens of thousands genes in brain space) [LHA+07] and microscale connectivity data (brain-network graphs with billions of edges [OHN+14, hum19, VESB+13, mPlDI+16]), whose utilization enables genetic dissection of brain networks on a genome-wide scale. Making these resources accessible for visual, quantitative, and qualitative exploration creates several challenges and research questions:

- **Missing analytical methods for multigenic brain functions/behavior:** *How do functionally associated genes relate to brain networks?* Brain function and behavior are both inherently multigenic. Consequently, identifying the neural networks through which these gene sets interact to express a given function or behavior is not trivial. Discovery tools that give computational predictions would provide an ideal entry point into this problem.
- **Size of the data:** *How can large collections of spatial brain data be explored?* Mining resources with tens of thousands of 3D images and voxel-level networks with billions of edges requires data handling techniques to make data with hundreds of gigabytes accessible without extensive computational power. While there are spatial indexing methods [BSG+09] that can be applied to 3D imaging data, there is a lack of such indices for voxel-level network data. Furthermore, visualization of these large networks is non-trivial in its complexity, therefore their exploration needs specialized visual analytics workflows.
- **Data on different scales and modalities:** *How can spatial data of different resolutions and modalities be related to each other?* Spatial data from different sources and of different modalities can vary in resolution, scale, and type [BS17]. Fusing, comparing, and exploring this data requires mapping to a common brain space, which depends on the task. For a global overview, interacting with the data on a brain region-level is sufficient, while small subnetworks for circuit dissection require a voxel-level resolution.

Motivated by these methodological gaps, we sought to investigate quantitative approaches to explore yet unknown functional neuroanatomy from genetic meta data that is underlying multigenic behavioral traits. This requires fusion of spatial connectivity data at varying scales, such as whole-brain correlated gene expression and structural and functional

connectivity with imaging and anatomical data. Most current analytical workflows in neuroscience involve only time-consuming manual data aggregation. Hence, visual analytics tools represent a promising approach to mine this data for insights into the functional organization of the brain [LGF+12]. Therefore, we envision new visual analytics methods for joint exploration and fusion of genetic data with multimodal brain network data on different anatomical levels. Bridging theses local-global scales, utilizing the hierarchical (e.g., from hierarchical brain parcellations) and spatial organization of the data is required, so it can be mapped to a common anatomical reference. On-demand queries on vast spatial data collections of gene expression and brain connectivity would then allow for interactive dissection of brain networks based on their spatial context. With ever-increasing resolution, this data exceeds the state-of-the art possibilities by several orders of magnitude in size and complexity. Incorporating techniques for handling big connectivity data is therefore a necessity. Taken together, these challenges lead to several overarching goals of this thesis:

- **G1) Workflows for joint exploration of connectivity with genetic data:** The combination of brain networks with gene expression data can lead to novel insights about functional neuroanatomy and behavior. Therefore, a central goal of this thesis is to develop analytical workflows to explore how genes relate to brain networks, and how these genes/networks relate to function/behavior.
- **G2) Handling large, multi-scale spatial connectivity and imaging data:** Accessing and querying large data sets in real-time is a key element for exploring them interactively. Data from different sources often have different resolutions and are seldom related to the same anatomical reference space. Mapping the data to a common reference space and utilizing fast data-accessing techniques therefore represent another central goal.
- **G3) Visualization and interaction techniques for heterogeneous big brain data:** Current neuroscientific workflows require time consuming manual aggregation of the data. Specialized visualization and interaction schemes for imaging, anatomical, and connectivity data can lead to more efficient workflows. These would enable neuroscientists to focus on understanding the data rather than managing it. Hence, they are of high relevance to this thesis.

1.2 Scope and Contributions

The underlying idea of this thesis is to combine neurological spatial data of different types - specifically 3D imaging data with brain networks to explore where genes have an effect on brain regions/networks and subsequently how they influence brain function/behavior. A first step in this quest was to research how genetic data has been previously analyzed in respect to brain networks, and how these results can be fused to map function/behavior (over genes) in the brain (G1). In Paper A, published in 2018 in *NeuroImage* [GKP+18],

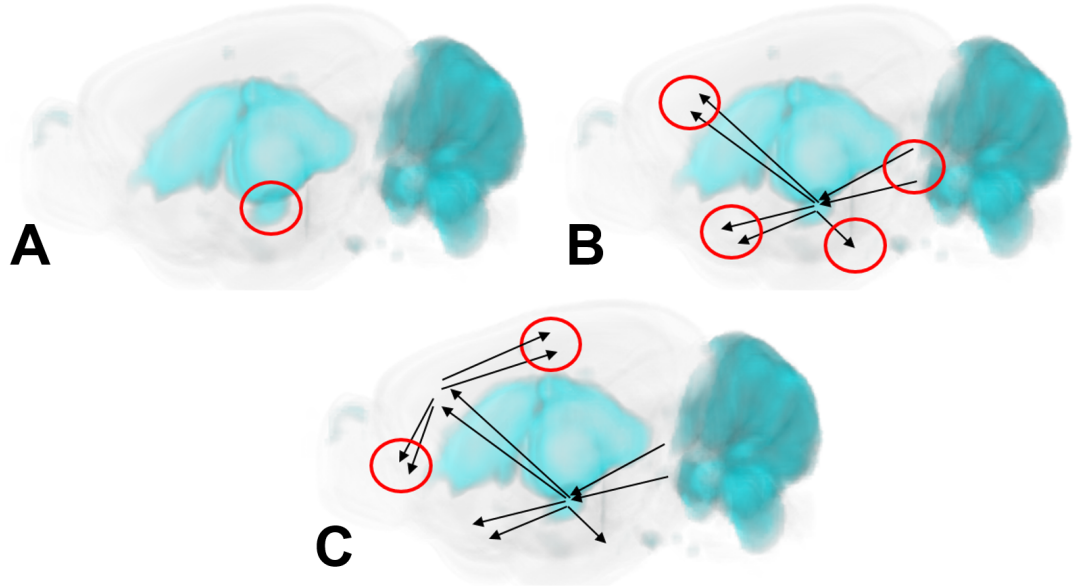


Figure 1.1: Effects of a gene in the mouse brain. Spatial gene expression of the gene PKC-Delta is shown in blue: (A) Primary expression sites: Effect on a brain region of interest where the gene is expressed (red circle). (B) Second-order effects: Projections from/to the brain region of interest (red circles). (C) Higher-order effects: Projections from/to the second-order effects (red circles).

a statistical method was proposed that is able to predict functional neuroanatomical maps (i.e., which brain areas are associated with a function) by introducing a genetically weighted connectivity analysis (GWCA). GWCA weights incoming/outgoing structural connections of areas where multiple genes have an effect, and computes the quantitative impact on the global structural connectivity with network measures. This procedure allows for high-throughput exploration of functional neuroanatomy or identifying behavioral or psychiatric traits associated with brain circuitry on a microscale-level in silico (G1). For an example, see Figure 1.2. These maps can then serve as entry points for neuronal circuit dissection in neuroscientific experimental workflows. We evaluated these maps based on known functional associations of genes in the mouse and human brain. A detailed discussion of the contribution can be found in Section 3.1. GWCA is script-based, tailored to structural brain connectivity, and involves extensive graph theoretical analysis. Hence, we sought to develop an interactive, more general framework that enables neuroscientists to fully utilize vast neurobiological resources - not limited by modality, size, and scale of spatial data. Since these resources involve thousands of 3D images and networks with billions of edges, interactive tools require specialized data-handling techniques to make continuous workflows possible (G2).

While methods for querying large collections of 3D imaging data exists, [BSG+09], there

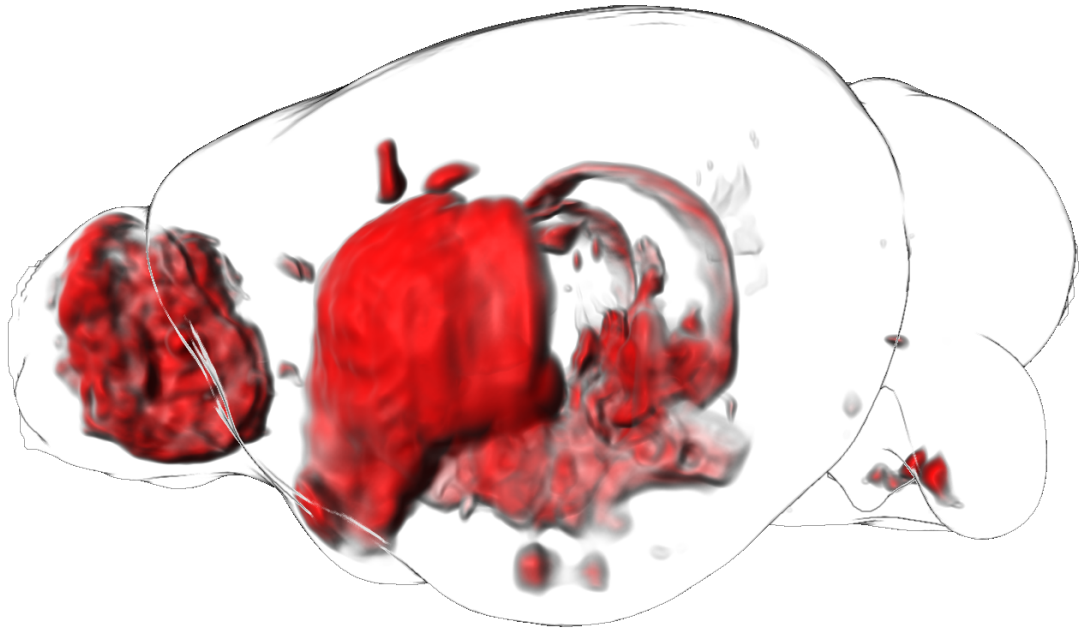


Figure 1.2: Predicted functional neuroanatomical map of social bonding behavior in the mouse brain. Red color indicates where the areas are and the transparency shows how strong these areas are associated.

is a lack of tools for quickly querying large spatial network graphs to make interactive applications possible (G2). Therefore, we created a data structure for real-time querying of big brain networks (Paper B), published in *NeuroInformatics* in 2019 [GKHB19]. The principle behind this data structure is the so called *Aggregation Query*, i.e., the aggregated connectivity from, to, or between volumes of interest (parts of the brain consisting of several voxels such as a brain region or a user-selected area). Hence, the data structure can provide connectivity on different levels, from single voxels to brain region-level (see Figure 1.3). This enables a comparison of brain networks independent of their scale and resolution. We demonstrated that even graphs reflecting tens of gigabytes of data results can be produced in an instant on consumer-level machines. This is achieved by harnessing spatial organization of the data, hierarchical brain parcellations, a tailored compression technique and caching. As a consequence, the data structure outperforms state-of-the-art graph engines by an order of magnitude. Furthermore, we showed in a prototypical web-component (G3) that *Aggregation Queries* can be used for inter-species comparisons of multimodal brain networks linked to autism - highlighting a potential use

in studying psychiatric conditions.

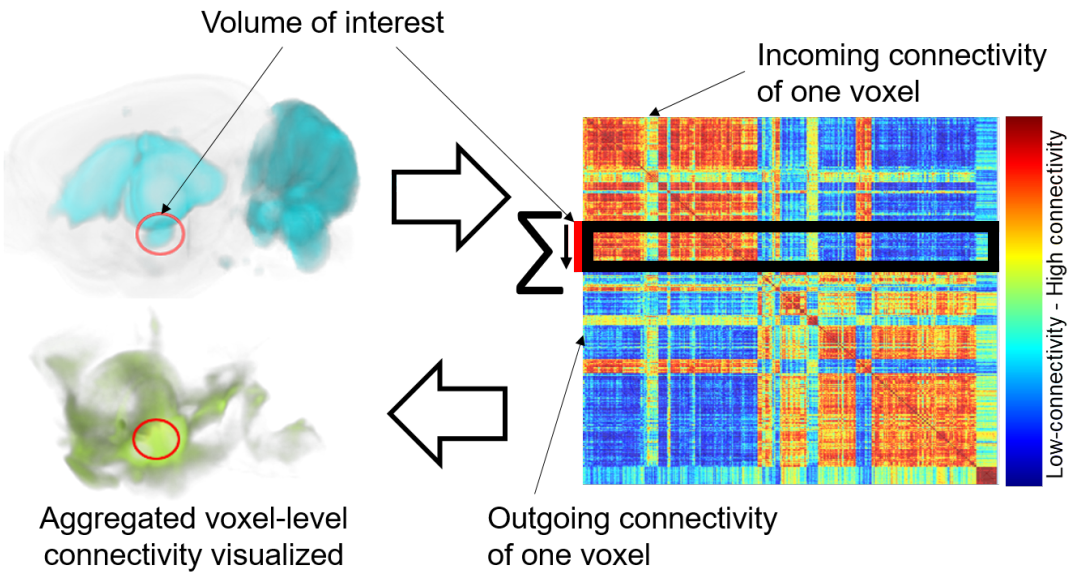


Figure 1.3: Aggregation Query: The aggregated outgoing connectivity of a brain region (red circle) can be computed by accumulating the outgoing connectivity of every voxel within the brain region in a connectivity matrix. The result is the sum of all outgoing connections from the brain region to the rest of the brain (green cloud).

With this data structure it was possible to create an interactive web-based, visual analytics framework called *BrainTrawler*. The first concept of this application has been presented at the *18th Eurographics Workshop on Visual Computing for Biology and Medicine* in 2018 (Paper C) [GSF+18] and has been extended to a complete version of *BrainTrawler* for *Computers and Graphics* in 2019 (Paper D) [GSF+19]. In these papers, we presented a tool that allows neuroscientists the exploration of big brain matrices on both a global and local level in real-time (Paper C and D) and also includes a genome-level spatial gene expression database (Paper D) that can be used to dissect networks genetically (G1), i.e., identifying which genes are active in different parts of a network. Connectivity data at different resolutions, such as mesoscale structural connectivity and region-wise functional connectivity, can be queried on different levels of a common hierarchical reference space. For an example, see Figure 1.4. This procedure allows neuroscientists to compare multimodal networks on different scales (G3). Additionally, 3D visualizations have been optimized to accommodate domain experts' needs for publishable network figures.

Together the four papers represent a diversity of steps and measures to allow neuroscientists the interactive visual analysis of heterogeneous spatial brain data. Figure 1.5 provides an overview on how the papers relate to the data and how they contributed to the effort of creating interactive visualizations. Paper A was the first step in analyzing

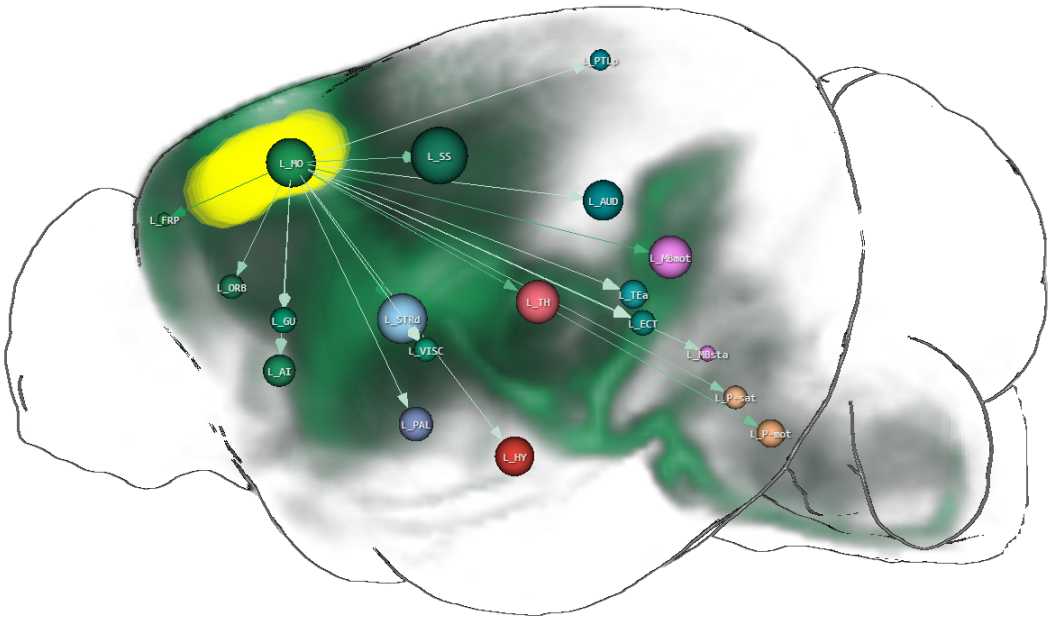


Figure 1.4: Exemplary sub-network in the mouse brain on different levels of anatomical abstraction: The green point cloud represents the outgoing structural connectivity of the yellow area. The 3D node-link diagram represents this connectivity at a brain region-level. Connectivity is outgoing from the left motor cortex (L_MO). Spheres identify the center of brain regions, while the size encodes the region size. The intensity of the arrows indicates the connection strength (i.e., how many voxels in this region have a connection from L_MO).

large scale connectivity matrices with spatial gene expression data. Here, the focus was on a method for quantitative analysis of the data and their visualization. It did not facilitate specialized data handling methods that would allow for interactive exploration of large scale data. For this purpose, we conceived the data structure in Paper B which provided a basis for more advanced visualizations. It directly led to Paper C and its extended version Paper D. The expertise we generated with domain experts during the development of Paper A was utilized to create interactive and more versatile quantitative data analysis workflows based on the principles of visual analytics. Thus, all papers build consecutively on each other, whereby data handling techniques enable increasingly interactive visualizations.

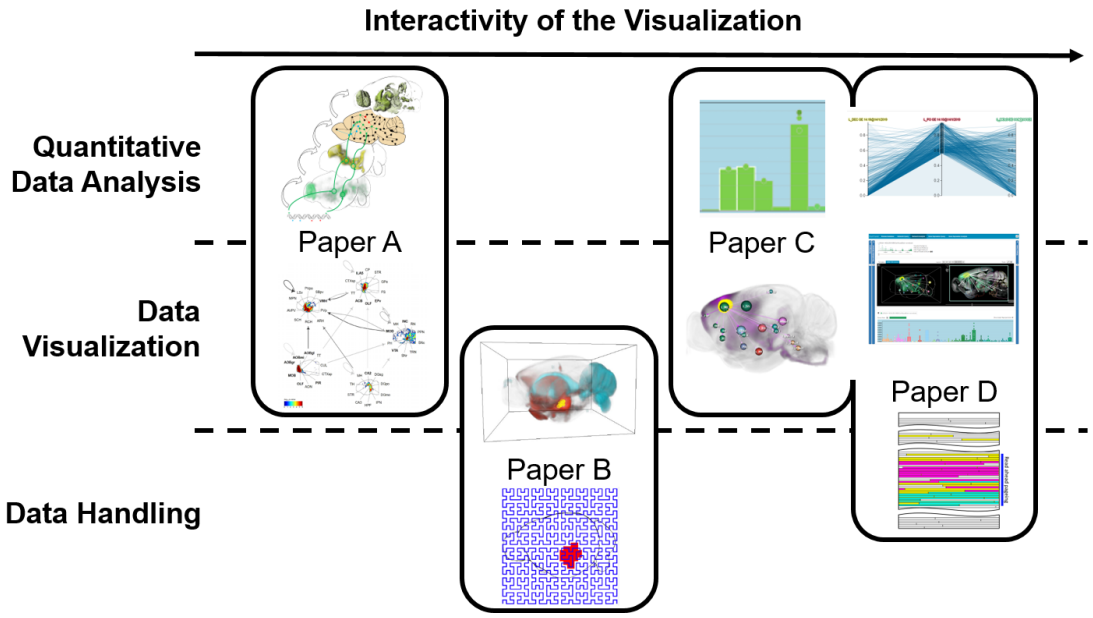


Figure 1.5: The papers of this thesis categorized by their level of interaction with the underlying data and the interactivity of the visualization. The figures indicate the papers' key contributions.

1.3 Contributions of Co-Authors

The papers of this thesis are a result of the author's research at the *VRVis Research Center* (*VRVIS*) in cooperation with the *Institute of Molecular Pathology Vienna* (*IMP*). The advisor, Eduard Gröller (*Technical University of Vienna*), was not involved as a co-author in these papers, but provided guidance and feedback.

Paper A (*NeuroImage* [GKP+18]) was jointly supervised by Wulf Haubensak, group leader of the *Haubensak Group* at the *IMP* and Katja Bühler, head of the *Biomedical Image Informatics Group* at *VRVis*, which were also involved in conceiving the method. The statistical and technical implementation, data preprocessing, data analysis, quantitative validation, writing the manuscript, and the the main effort in conceiving the method was done by the author of this thesis. The paper was co-authored by Joanna Kaczanowska (*IMP*, *HaubensakGroup*) who performed the qualitative, neurobiological validation and took also part in conceiving the method. Wulf Haubensak, Katja Bühler and Joanna Kaczanowska took also part in a joint effort to rewrite and polish the text for the target audience of *NeuroImage*. Josef M. Penninger (*Institute of Molecular Biotechnology of the Austrian Academy of Sciences*) and Andreas Hess (*Institute of Experimental and Clinical Pharmacology and Toxicology, Friedrich-Alexander University Erlangen-Nuremberg*) provided the fMRI data and behavior-associated gene sets.

Paper B (NeuroInformatics 2019 [GKHB19]) was supervised by Katja Bühler, who supported writing the manuscript and provided mathematical expertise. The author of this thesis conceived the method, implemented the data structure, performed quantitative and qualitative evaluation, wrote the manuscript, and created the major parts of the web-component. Wulf Haubensak and Joanna Kaczanowska provided neurobiological expertise and helped designing the case studies for evaluation. Florian Schulze, Nicolas Swoboda, Markus Töpfer, and Emre Tosun (all current or former members of the *Biomedical Image Informatics Group* at *VRVis*) were involved in creating parts of the web-component that has been used for the case studies.

For the Papers C and D, the web-component of Paper B was upgraded, adapted, reused, and refined to be presented at the *18th Eurographics Workshop on Visual Computing for Biology and Medicine* [GSF+18] and its extended version, *BrainTrawler*, was published in *Computers and Graphics* [GSF+19]. The papers were again supervised by Katja Bühler, who provided valuable advice and supported the writing of the manuscript. The author of this thesis designed and implemented the framework, created the data structures, conducted the case studies, performed the evaluation, and wrote the manuscript. Neurobiological expertise and help with the case studies was given by Wulf Haubensak and Joanna Kaczanowska. Nicolas Swoboda created the silhouette visualization of 3D brains. Lisa Frauenstein, a former master student at *VRVis* assisted with literature research for the related work section.

1.4 Thesis Structure

The first part of this thesis (Chapters 1 - 4) describes the individual contributions of the papers and how they relate to each other in a bigger picture. Chapter 2 presents an overview on spatial brain data and describes in detail the types of data that have been used in this thesis. Furthermore, it outlines the state-of-the-art visual analytics tools used with this kind of data. Chapter 3 highlights the papers' individual contributions. Chapter 4 discusses conclusions, the impact of this thesis on the field and collaborators, and an outlook of future work. The second part of this thesis consists of the published papers itself i.e., Paper A, B, C and D.

CHAPTER 2

Background and Related Work

A central aim, from basic neuroscience to psychiatry, is to resolve how genes control brain circuitry and behavior [SKI+17]. This is experimentally hard, since most brain functions and behaviors are controlled by multiple genes [HKC+10, MN08]. Figure 2.1 outlines how genes-brain-behavior relationships can be determined. Genetic alteration(s) in an animal model (e.g., mouse) can be used to study the behavior/function of genes, while fMRI (functional magnetic resonance imaging), electrophysiology (measurement of electrical activity of neurons), or optophysiology (visualizing cellular activity with fluorescence microscopy) can subsequently determine which brain regions/networks are involved (Figure 2.1, black arrows). This requires breeding of genetically modified animals that perform cognitive tasks while their neuronal activity is recorded [HLM+15]. In this low throughput, it is difficult to delineate the neural circuitry through which these sets of genes express their behavioral effects. To capture brain regions/networks that rely on multigenic behavior/function (i.e., dependent on multiple genes), the increasing amount of publicly available brain and genetic data offers a rich source that can be mined to address this task computationally. This requires a joint exploration of behavior-associated genes from literature, genetic databases or GWAS (genome-wide association studies), and spatial brain data (3D images of gene expressions, spatial networks, and hierarchical brain parcellations) targeted by statistical and visual analytics methods (Figure 2.1, red arrows).

In recent years, the role of visual analytics in neuroscience has become increasingly important with the emergence of high-throughput imaging techniques. These techniques have created a wealth of resources for which data mining requires manual data aggregation via scripting, and consequently the expertise of a bioinformatician. Visual analytics tools bridge these gap by enabling neuroscientists to interactively browse vast data collections, visualize complex relationships, and link different types of data. A seamless exploration without intermediate data analysis by statisticians/computer-scientists/mathematicians

creates continuous workflows that can support the process of analytic reasoning in a neuroscientific environment.

This chapter discusses the context of visual analytics methods for exploring spatial data in circuit neuroscience. An overview of spatial brain data in the context of genes-brain-behavior relationships is presented in Section 2.1. Section 2.2 provides a state-of-the-art report about current visual analytics tools that handle and fuse these data. It especially focuses on the interactive exploration of brain networks, since network analysis is crucial for understanding the interactions of neurobiological systems [BS17].

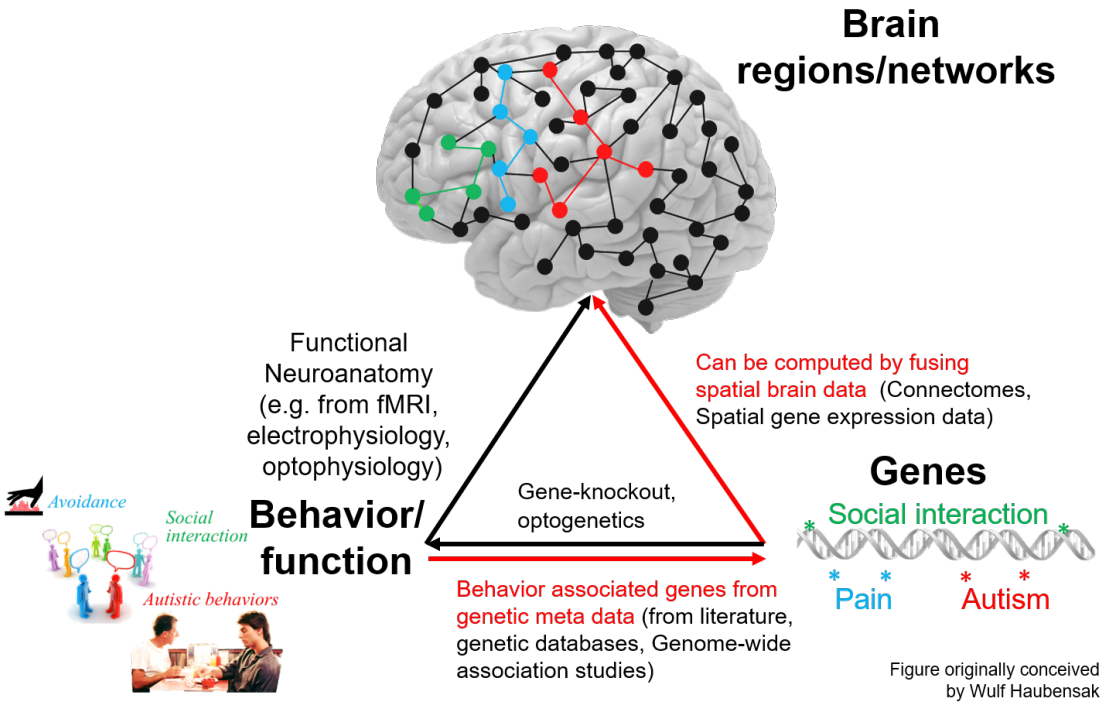


Figure 2.1: Determination of relationships between brain regions/networks, behavior/functions and genes. Black arrows: Alterations in the genome of model organisms (e.g., mouse) can be studied to identify the function of specific genes and their influence on the organism's behavior. Via fMRI, electrophysiology, or optophysiology, these effects can be linked to brain regions or networks. Red arrows: From literature, genetic databases, or genome-wide association studies, collections of genes can be associated to behavior/function. By mapping these genes to the brain via spatial brain data of big brain initiatives, the effect of behavior/function/genes on the brain can be studied computationally.

2.1 Spatial Brain Data to Explore the Relationship between Genes, Brain, and Behavior

This section summarizes spatial data that is used to explore the genes-brain-behavior relationship in neuroscience. For this purpose, basically three types of spatial data are relevant: Imaging data that relates function/behavior/genes to spatial location in the brain, connectivity data to provide relational information between these locations, and anatomical data for spatial context (e.g., anatomical annotations and brain parcellations). Note, that although genetic meta data such as behavioral/functional annotations of genes play a major role, they are not inherently spatial. Therefore, they are not included in this enumeration.

2.1.1 Imaging Data

In circuit neuroscience, spatial imaging data is used to relate a structural, functional, or genetic property to spatial locations. This allows researchers to draw conclusions about which brain regions contribute to a behavior, which function a certain brain region has, and which molecular mechanisms might be involved (from genetic data). One can distinguish these spatial locations on two different anatomical levels: on a brain region-level, where the data is associated to anatomical (e.g., thalamus, cerebellum) or non-anatomical (e.g., electrode positions) brain regions, and on a voxel-level, where voxels refer to grid points in a regular 3D space on a sub-brain region-level (Figure 2.2B).

Region-level imaging data represents measurements that are only available for brain regions (i.e., one data point for each region). The data is generated via imaging techniques for which voxel-level resolution is inherently impossible (e.g., data from probes on the scalp) or not feasible (e.g., data is from biopsy-sites). For example, functional near-infrared spectroscopy (fNIRS) data measures the hemodynamic response via optical sensors placed on the scalp [FGSZ85]. The hemodynamic response shows where the blood flow in the brain is localized and therefore indicates neural activity. Another probe-based imaging technique represents electrophysiology, which is used to measure voltage changes resulting from electrical currents within neurons. Non-invasive, on a brain-wide scale, electroencephalography (EEG) offers high temporal resolution, so it can be used to measure event/task related activity [SdS12]. The electrodes can also be implanted into the brain to offer subcortical recordings, which is then called electrocorticography (ECoG). On a cellular/neuron scale, electrophysiology can be used to measure intracellular/extracellular action potentials in neurons (e.g., via microscopic clamps). These “nerve impulses“ can be recorded with a temporal resolution in the range of milliseconds and can therefore be used to study neuronal circuit dynamics (how brain regions communicate) or neurotransmission [SH09]. A different way to generate region-level imaging data are biopsies. Since this involves taking living cells from the brain, biopsies are usually done for tumor sectioning or post-mortem. For example, for the Allen Human Brain Atlas, Hawrylycz et al. [HLGB$^+$12] took samples from 900 neuroanatomical subdivisions of the brain (i.e., brain regions) from several donors and

performed a microarray analysis to measure the gene expression levels on a genome-wide scale. With this data it is possible to identify where genes are expressed in the brain, and, therefore, to draw conclusions regarding their function.

Voxel-level imaging data is volumetric data generated with 3D data acquisition techniques. The data is divided by a 3D grid into voxels that represent measurements at their respective positions (i.e., one data point for each grid position/voxel). In recent years, a variety of different 3D neuroimaging techniques have been developed, of which several are relevant for neurocircuit research. Functional magnetic resonance imaging (fMRI) represents one of the most prominent tools in neuroscience [Pol08] as it is non-invasive. Similar to fNIRS, it measures the hemodynamic response in the brain, but with higher spatial resolution where voxels are in the range of millimeters. Magnetocephalography (MEG) is often used complementary to fMRI, because it has a higher temporal resolution ($\sim$ 1 millisecond compared to $\sim$ 1 second) but a lower spatial contrast. Other methods that operate on this level of detail are single photon emission computed tomography (SPECT) and positron-emission tomography (PET) [KLVV13]. They enable a functional mapping of metabolic activity for which they are tailored to investigate neurological diseases [LY15]. On a smaller scale, optophysiological techniques made it possible to observe the activity or genetic properties of single cells/neurons [SMT13]. Advances in confocal and fluorescence microscopy enable unprecedented spatial resolution of neuronal activity (calcium imaging [SGHK03]), neuronal structure, and gene expression (genetically encoded fluorescence proteins [NRR$^+$10]) in living animals. Applied in high-throughput, large image collections can be generated. One particular example represents the Allen Mouse Brain Atlas [LHA$^+$07], which provides spatial gene expression for the mouse brain on a genome-wide scale. This shows, where in the brain is which gene expressed on a 200 microns resolution. Hence, it enables the genetic dissection of the brain on a voxel-level.

2.1.2 Anatomical Data

Anatomical data is necessary to relate imaging data to its anatomical context. They are not a single type of data, they rather represent a diverse collection of reference templates, anatomical parcellations, and neuroanatomical ontologies. Together they form the common knowledge of how the brain is structured and how this structure can be referred to. For example, the prefrontal cortex lies in the front part of the human brain and is linked to complex cognitive behavior [YR09]).

A reference template is structural imaging data that has been combined (e.g., via image registration) to a structural representation of the brain for a group of specimen or a species. It can be used as common reference space for imaging data to enable voxel-level correspondence for a collection of images, like the Allen Mouse Brain Atlas [LHA$^+$07] (Figure 2.3A). Moreover it provides spatial orientation when visualized with other spatial data, see for example the gene expression in Figure 1.1.

A neuroanatomical ontology is the formal representation of knowledge about the anatomy of the brain [LM09] of a species. This relates foremost to the composition of the

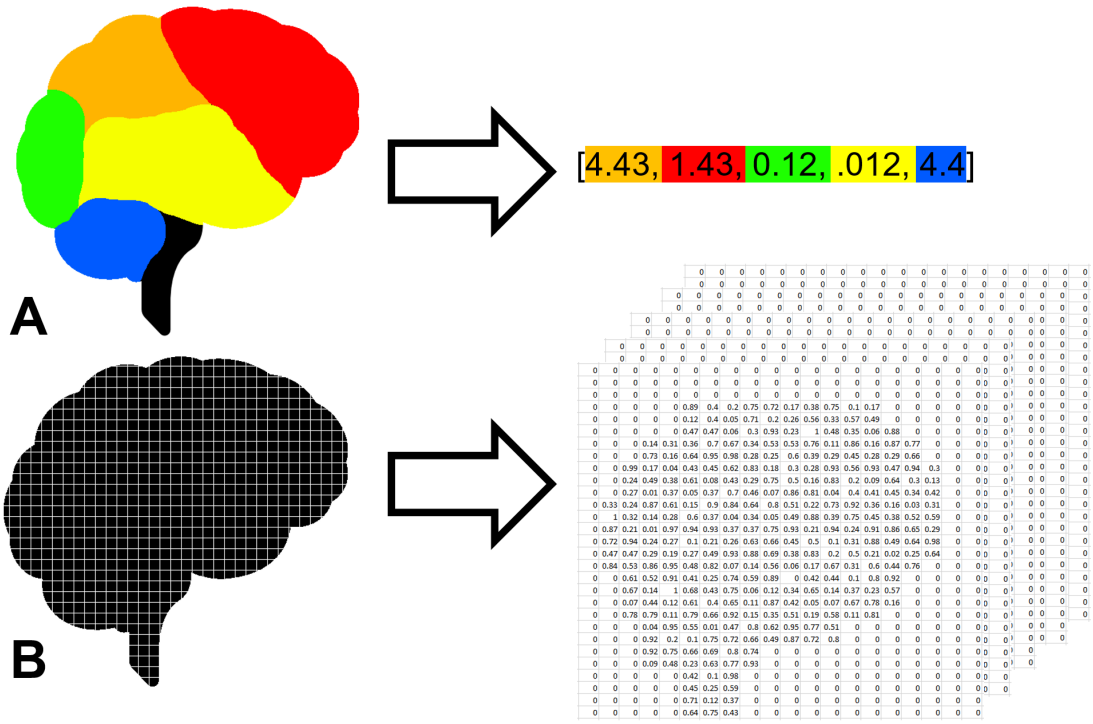


Figure 2.2: Imaging Data: (A) Region-level data: One value for every brain region. (B) Voxel-level data: Volumetric data, one value for every voxel.

brain, i.e., which brain regions it consists of and how these brain regions are subdivided (hierarchically). It may also include naming or color conventions. Figure 2.3B shows a section of the Allen Reference Atlas [LHA+07].

Anatomical parcellations act as links between neuroanatomical ontologies and reference templates. In principle, it consists of a regional annotation of every voxel in a reference template, or a continuous outline around a region (e.g., a mesh). Hence, voxels can be associated with brain regions of an ontology for visualizing anatomical context (Figure 2.3C) and relating voxel-level to region-level data.

2.1.3 Connectivity Data

Connectivity data represent the relations between different spatial locations in the brain of a certain modality. Connectivity is organized in weighted adjacency matrices, so-called connectivity matrices. Columns/rows represent the incoming/outgoing connectivity between brain areas on either voxel- or region-level. In the field of network neuroscience, there is no general consent of how to fundamentally distinguish types of connectivity data. Olaf Sporn [Spo13] divided connectivity into anatomical/structural connectivity (anatomical links), functional connectivity (statistical functional dependencies), and effective connectivity (directed causal effects). Betzel and Basset [BB17] referred to

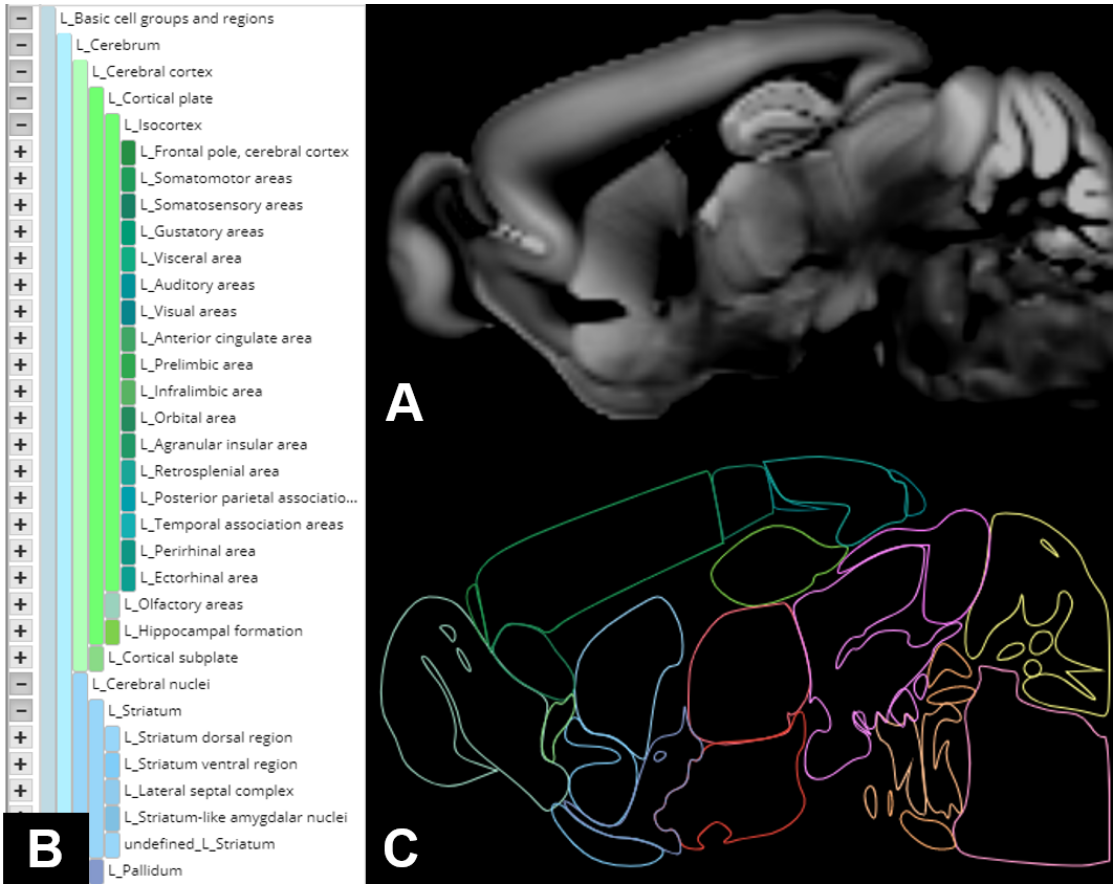


Figure 2.3: Anatomical Data: (A) Anatomical reference template showing the anatomical structure of a mouse brain (2D sagittal slice). (B) Part of a hierarchical neuroanatomical ontology. (C) Brain parcellation that corresponds to the brain ontology (2D sagittal slice).

structural and functional connectivity, but omitted effective connectivity. Cauda et al. [DNM+18] used a similar discrimination, but added genetic connectivity, describing the correlation of spatial gene expression (i.e., if brain regions express similar genes). In the following enumeration, we included all of the mentioned data types, because they represent non-overlapping distinct modalities.

Anatomical/structural connectivity describes how brain areas are physically connected via neuronal projections (axons). These projections are on a scale that recording them is performed on a voxel-level, but depending on the acquisition techniques (i.e., noise level, spatial resolution) they may be aggregated to region-level information [BB17]. The connectivity is usually sparse, for single neurons are usually not connected to the entire brain, but rather to distinct regions [Spo13] (Figure 2.4A). One example for structural connectivity is DTI (diffusion tensor imaging), which uses MRI to map the diffusion pro-

cess of molecules in biological tissue. It measures the diffusion tensor - the directionality of the diffusion of water - to estimate fiber tracts of neuronal projections on a voxel-level. Via tractography, these tracts can be traced between brain regions to generate a region-level [Laz10, CBC+15] white matter (anatomical) connectivity matrix. Recent advances in high-throughput microscopy enabled the generation of anatomical connectivity based on neuron bundles. Oh et al. [OHN+14] created voxel-level connectivity at a resolution of 100 microns by injecting viral tracers into the mouse brain. Hence, a directed connectivity matrix can be generated that maps efferent neurons from $\sim 15\%$ of the brain. Even higher resolution on a cellular level can be achieved by using electron microscopy data. For example, Zheng et al. [ZLP+18] published a complete electron microscopy volume of the drosophila melanogaster brain that allows researchers to create connectivity for individual neurons provided that these neurons can be accurately segmented.

Functional Connectivity represents the statistical dependence of brain areas during task performance or a sensory stimuli [Spo13]. Hence, it reflects how brain regions functionally synergize. Functional connectivity is usually denser than anatomical connectivity, as it contains connections to anatomically unconnected regions [RS10] (Figure 2.4B). The data can be derived from time series observations, i.e., imaging data that has been recorded for discrete time points over a period of time. During these recordings, sensory stimuli are applied or tasks are performed. The cross-correlation of voxels/brain regions over time reflects the functional relationships and therefore (undirected) functional connectivity. This requires data acquisition with a high temporal resolution of multiple brain regions simultaneously, such as with fNIRS, EEG, MEG, and fMRI. These methods may suffer from noise and signal dropout as well from computational challenges regarding a large number of voxels, so it is common to analyze functional connectivity on a region-level rather than voxel-level [dRvdH13].

Effective Connectivity describes the causality of interactions, so it can be seen as the directionality of functional relations [Spo13]. It cannot be computed by a single acquisition technique, rather it can be derived from a combination of structural and functional connectivity. While functional connectivity is undirected, the flow of information can be inferred from the directed structural connectivity [HKBS07]. Hence, it represents a weighted directed functional connectivity matrix (Figure 2.4C). Another possibility is the perturbation of brain regions via transcranial magnetic stimulation or invasive methods such as deep brain stimulation. The resulting changes in functional connectivity can then be used to model causal relations [MFH+05].

Genetic Connectivity quantifies the transcriptional similarities between brain areas (i.e., how similar are they in their molecular mechanisms) [DNM+18]. The availability of spatial gene expression on a genome-wide scale [LHA+07, HLGB+12] made the computation of gene co-expression correlation across the whole brain possible[RA15]. For this, the correlation of the expression of genes is calculated between the voxels/regions in the brain. This gene co-expression correlation matrix, or (undirected) genetic connectivity (Figure 2.4B) makes it possible to explore the genetic mechanisms in the brain for function/behavior/disease related sets of genes.

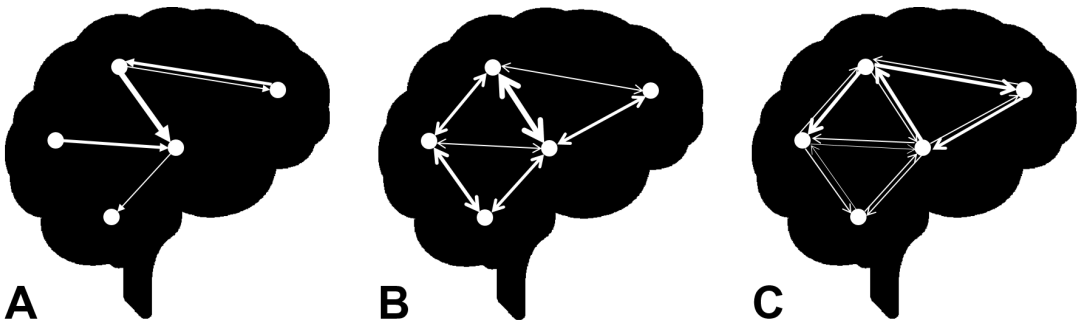


Figure 2.4: Connectivity data types: (A) Sparse directed network (structural connectivity) (B) Dense undirected network (functional connectivity, genetic connectivity) (C) Dense directed network (effective connectivity).

2.2 Visual Analytics Tools for Spatial Brain Data

Recent studies have shown that the combination of imaging, anatomical, and connectivity data can lead to novel insights into the neurocircuitry and a better understanding of the brain's organization [FP11, JFD14, RA15, RYWB15, WVRG+16, FTP11, FF16]. Especially connectivity data represents a common modality visualized in many neuroscientific applications [MBWG13]. Therefore, this section is devoted primarily to the interactive exploration of brain connectivity. Section 2.2.1 is about the exploration of brain networks in relation to their anatomical context, Section 2.2.2 describes methods for comparing them. Section 2.2.3 gives an overview on tools for the exploration of spatial big brain data with respect to large-scale image databases and massive volumetric data.

2.2.1 Exploration of Brain Networks with respect to Anatomical Context

In recent years, an abundance of toolboxes have been published [RS10, GDL+11, RLF15] that offer computation and visualization of multimodal connectivity data. While they provide a rich set of statistical and mathematical methods, their visualizations are static and they often require experience in Matlab/Python scripting. In contrast, visual analytics tools support the processing of complex information via interactive visualizations, so neuroscientists can focus on understanding the data rather than handling it. This section gives an overview on visual analytics tools for analyzing brain connectivity data. First, general approaches for exploring brain connectivity data in a 3D anatomical context are described. The second part consists of tools that focus on a 2D abstract visualization using anatomical or intrinsic graph layouts. Note that tools that are intended for connectivity analysis on big brain data (e.g., BrainExplorer [FLN+15], CATMAID [SCHT09], BrainGazer [BSG+09], etc.) are presented in Section 2.2.3 and are therefore not part of this section.

A common way to visualize brain networks in neuroscientific publications are 3D node-link diagrams [ZFB10, RA15, BS17]. In these diagrams, network connections (edges) are often rendered as straight lines or arrows between spheres representing brain regions (nodes) across a 3D anatomical representation of the brain to help neuroscientists to orient themselves (Figure 2.5A). The BrainNet Viewer [XWH13] uses this type of visualization to depict region-level functional connectivity of experimental data in the human brain. Nodes and edges can be colored to encode network measures, edge weights, or different kind of additional information. Networks can be displayed in multiple views that allow rudimentary user interactions to display more detailed information (for example, clicking on the brain surface to get labels or measurements. A similar approach is used by the Connectome Visualization Utility [LDTS14]. In addition to the node-link model, it offers a matrix (heatmap) and a circular representation (i.e., a connectogram [ICT+12]) of the network in separate views that are linked with each other. Hence, interactions in one view are simultaneously performed in other views. These views offer a selection/highlighting of nodes and edges, so one can focus on specific parts of complex networks. Bezgin et al. [BRSK09] also employed user-selected nodes to visualize only relevant subnetworks in the Macaque monkey brain. In this case, brain regions from a hierarchical ontology can be chosen to define which connections should be shown as arrows overlaying 3D brain anatomy (i.e., a 3D node-link diagram without depicting the nodes).

3D node-link diagrams are also commonly used for simulation data. Nowke et al. [NSvA+13] introduced VisNEST, a tool that integrates macroscopic structural connectivity data of 32 brain regions with microscale simulated neural activity of the Macaque monkey's visual cortex. Connectivity is presented as a 3D node-link diagram, where the link thickness encodes connection strength. Anatomical context is provided with semi-transparent mesh renderings of the anatomical brain regions, visualized in parallel with the connectivity. In addition, simulation-related information is presented in a "population" view, depicting simulated neuronal populations, and a "flux" view visualizing time-varying activity across regions. A different approach for 3D network visualization has been proposed by Schmitt et al. [SE12] with neuroVIISAS, a tool for the multiscale simulation of neurons in the rat brain. NeuroVIISAS does not require a specific type of connectivity, rather it can be set manually, imported from a file, or generated randomly for testing. For network visualization, regions are not rendered as spheres, they are replaced by anatomical surface meshes, color coded based on a reference ontology. Regions can be selected from a hierarchical parcellation, so the hierarchical level for the analysis can be chosen interactively. This selection is linked with different 2D representations of the network, such as a hierarchical connectivity matrix and node-link diagrams with different layouts. Furthermore, textual queries can be performed to filter connections by their value or corresponding brain region.

With an increasing number of connections, i.e., in dense voxel-level connectivity data, a 3D node-link diagram produces clutter and obscures the anatomical context. Böttger et al. [BSJ+14] targeted this problem with a dual approach for "connexel" visualization (connectivity between voxels, i.e., voxel-level connectivity). For this purpose, they adapted

brainGL [bra11], an open source software for the interactive exploration of structural and functional brain data with edge-bundling (Figure 2.5B). By grouping geometrically similar edges to bundles, the overall structure of the network can be clarified. To visualize the termination points of these edges, various types of glyphs can be projected to the surface of the brain. The glyphs provide a visual summary of all connections from their respective position, which can be interactively manipulated by the user. Similar to the Connectome Visualization Utility, visualizations can be viewed in multiple linked views with synchronized selections, thresholds, and coloring.

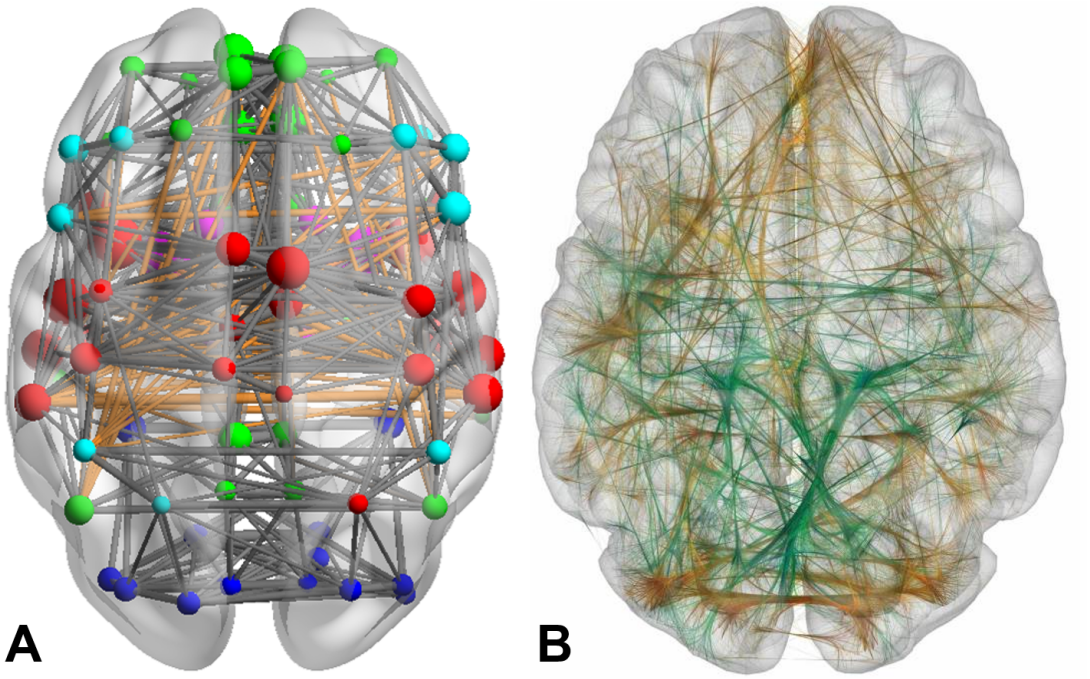


Figure 2.5: 3D network visualization. (A) Region-level network visualization via 3D node-link diagram in the BrainNet Viewer ([XWH13], Figure 5). Colors represent different subnetworks, node size represents node strength (number of connections of a node). Orange edges represent long distance connections. (B) Voxel-level network visualization of functional connectivity (orange: negative correlation, green: positive correlation) with edge-bundling ([BSJ+14], Figure 13).

Although the 3D spatial representation of networks provides anatomical context, 2D node-link diagrams with flexible layouts are better suited for comparing connectivity[ABH+13] or identifying modules (well-connected groups of nodes) [PLK+15]. For this reason, Brain-Modulizer [MBB+16] uses a linked presentation of anatomy in 3D, and network graphs in 2D to enable neuroscientists to interactively explore functional connectivity. Spatial correspondence is indicated via color coding (Figure 2.6) of hierarchically organized brain modules, but can be also established via brushing/selecting nodes in one of the views. Murugesan et al. [MBB+16] showed that with force-directed layouts it was possible to

perform graph theoretical analyses for up to 200 brain regions. Analogous to BrainModulizer, BRAINtrinsic [CYF+15, CYA+16] aimed to explore brain connectivity with node-link diagrams based on network topology. Instead of arranging nodes, they mapped the network to a topological space by taking the networks intrinsic geometry into account. For this purpose, they performed dimensionality reduction (multidimensional scaling, isomap, and t-distributed stochastic neighbor embedding) on structural and functional connectivity data. In a 3D view that shows the network as a node-link diagram, one can interactively switch between anatomical and topological spaces, show/hide particular brain regions and compute network measures. This approach has been taken further in the NeuroCave visualization system [KZA+18], optimized for virtual reality environments. Networks are shown in a linked side-by-side rendering (Figure 2.7), so the network is visible in both a 3D anatomical space and a topological space simultaneously. This combines the advantage of 3D spatial representations with the flexibility of node-link diagram layouts. Additionally, NeuroCave supports several methods to interactively change the network appearance, such as edge-bundling, color schemes, clustering, thresholding, and labeling.

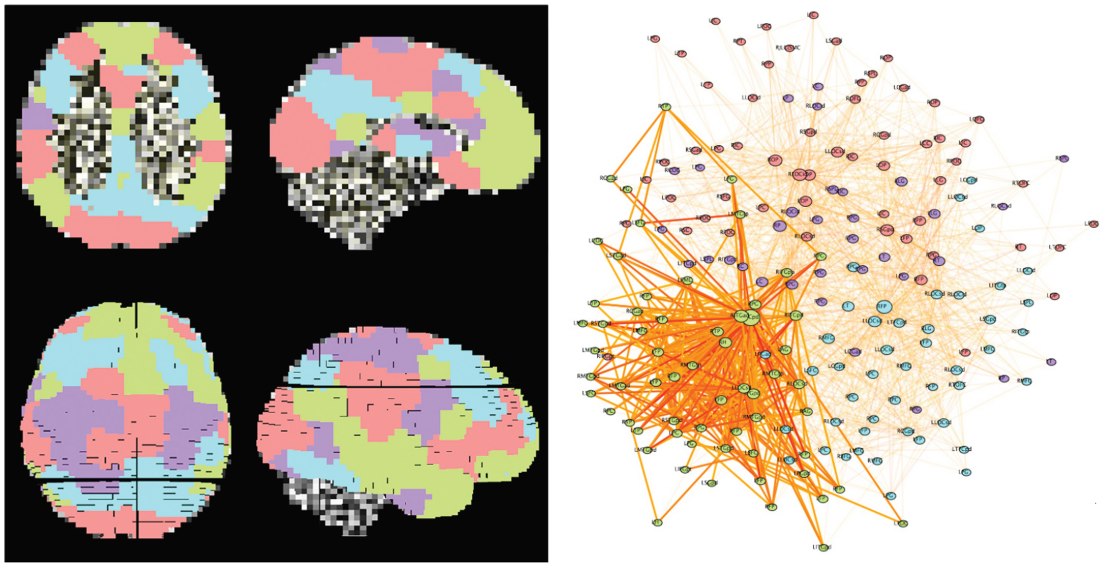


Figure 2.6: Node-link diagram for network visualization with anatomical context. BrainModulizer ([MBB+16], Figure 8) shows a 2D graph, split into four brain modules (red, blue, green, purple) rendered with a force-directed layout (B) next to its 3D anatomical context (A, colors correspond to nodes).

As an alternative to visualize the anatomical context in addition to node-link diagrams, the context can be also integrated directly into the graph layout. These so-called anatomical layouts are abstract 2D representations of brain regions, i.e., the 3D brain anatomy is flatted to a 2D space. NeuroMap [Sor13] uses anatomical layouts to map

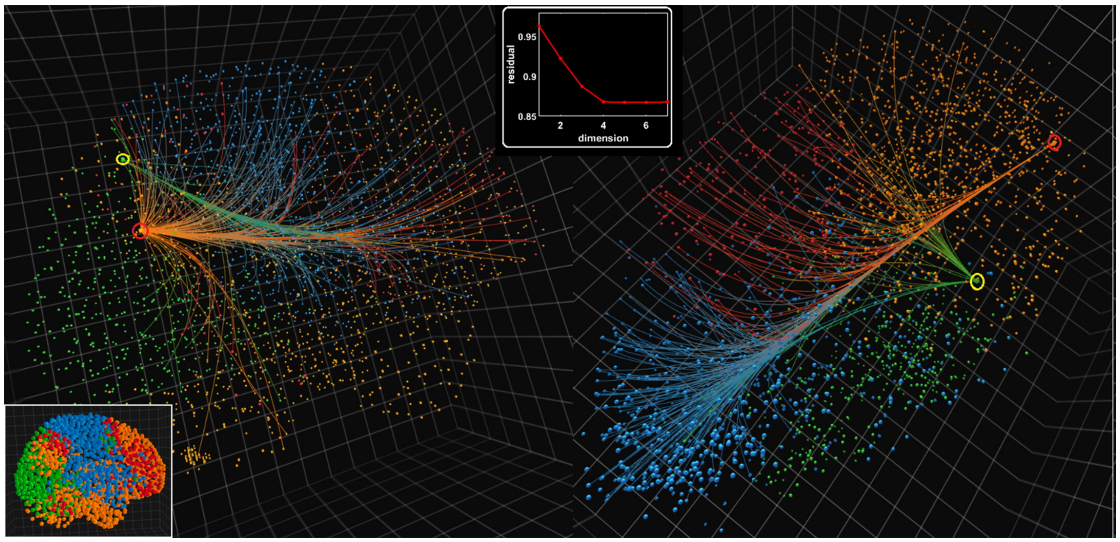


Figure 2.7: A network graph in 3D anatomical space (left) showing four brain modules (orange, blue, green and red) visualized in NeuroCave ([KZA+18], Figure 5). The right panel shows the same graph, visualized in a topological space (connectivity-driven).

potential neuronal circuits in a fruit fly's brain as interactive wiring diagrams. For this purpose, fixed compartment positions (Figure 2.8A) that have been manually defined in collaboration with neuroscientists are used to depict the overall structure of the brain. The visualization can be interactively adapted by adding new connections from additional data, filtering, highlighting, or layout adjustments. A similar, static, visualization approach has been used by Ji et al. [JMR18], which maps functional networks derived from EEG to a planar projection of the human skull. To avoid cluttering, only the connectivity of one functional unit (i.e., network module) can be shown in a single image (Figure 2.8B). Although this tool does not represent a visual analytics approach, it is included in this section because its relevance for anatomical layouts.

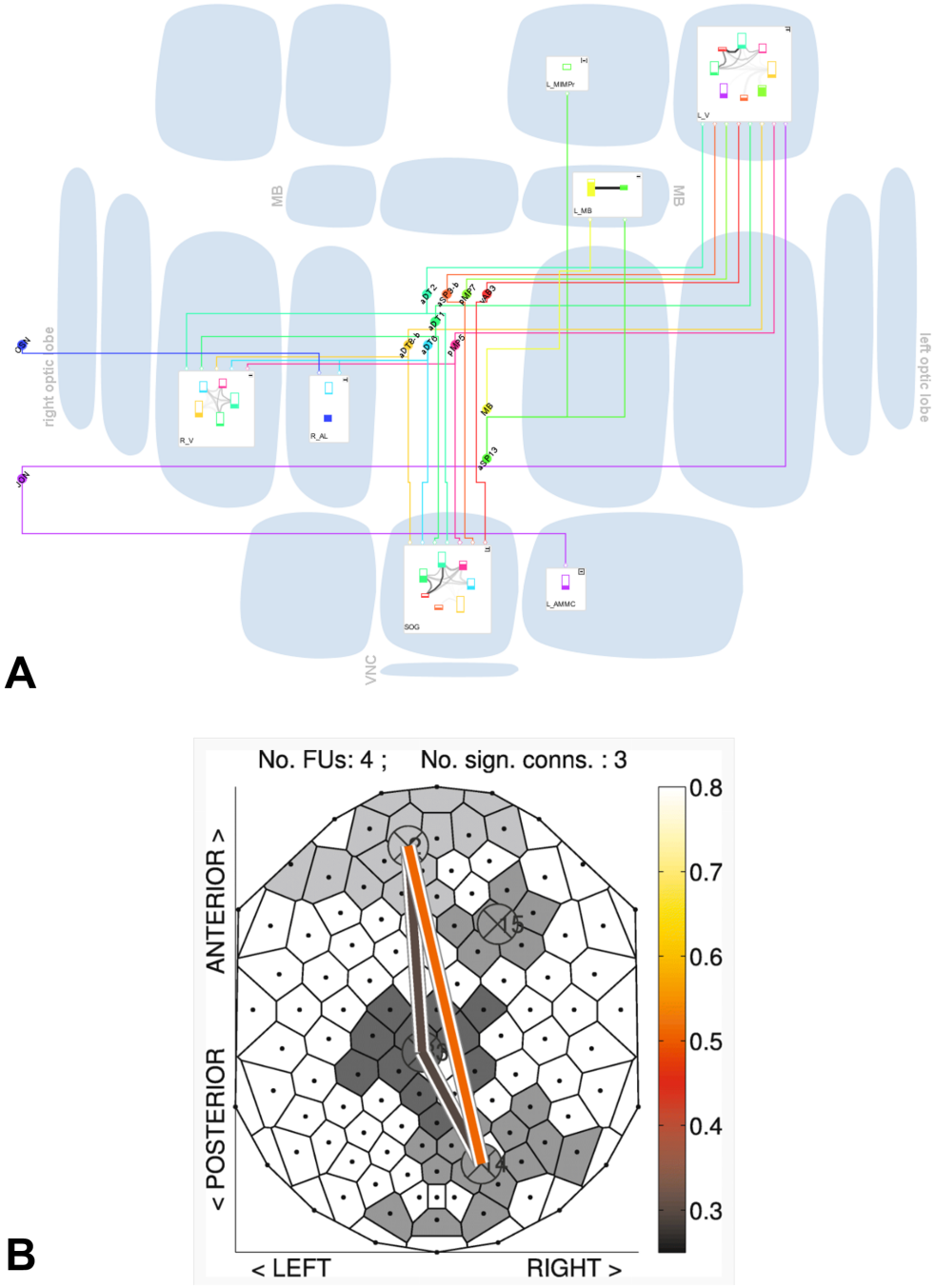


Figure 2.8: Anatomical layouts: (A) Anatomical layout of a fruit fly brain (drosophila). Blue regions are major brain regions of the fruit fly brain. Network nodes are arranged within these regions (NeuroMap [Sor13], Figure 6). (B) Planar projection of 3D electrode locations (nodes) on the human skull. Functional connectivity of a network module is visualized between these locations (Ji et al. [JMR18], Figure 7).

2.2.2 Brain Network Comparison

Visualizing the similarities/differences of networks is essential for comparing different kinds of connectivities [ABH+13] or tracing network changes over time [MKF+15]. Alper et al. [ABH+13] conducted a controlled design study, which suggested that matrix visualization in combination with glyphs outperforms superimposed node-link diagrams of two different brain connectivities. To combine the node-link diagram's ability to simplify network module identification with matrix visualization, Ma et al. [MKF+15] proposed a dual-representation to explore dynamic functional networks. In this approach, network changes can be traced via animations between network transitions - the change of a network state from one time step to the next - while the change from the previous transition is visualized in a matrix with glyphs. A way to do this without animation are small multiples, a series of similar graphs with the same scale to compare them easily. De Ridder et al. [dRKY+18] used this method to render network graphs in a circular layout with similar anatomical regions/nodes to compare functional connectivity (Figure 2.9A). Spatial context is provided via a linked 3D anatomy viewer, which can be used to select brain regions that are shown in the small multiples. Selections in the small multiples are highlighted in the anatomy viewer, which can be used to trace uncertainty in fMRI data, for example head movement that leads to an unnatural high connectivity between spatially close regions. This method has been also employed by BRAVIZ [ASO+16]. It is a rather general tool that allows neuroscientists the analysis of human fMRI and DTI image data in combination with an anatomical reference space. BRAVIZ is tailored to a variety of workflows, such as DTI fiber tractography, functional fMRI analysis, or validating image registration. Although it does not explicitly facilitate network visualization or analysis, it enables the comparison of DTI fiber bundles via small multiples, and therefore indirectly, of structural connectivity.

For dynamic networks with hundreds of time points, animations or small multiples become increasingly time-consuming and unreliable as they rely on memorization by the user. Therefore, Bach et al. introduced Small MultiPiles [BHRD+15] to identify temporal patterns in functional networks on such long time scales. They used a piling metaphor to visually encode snapshots of a network (i.e., a connectivity matrix at each time point) into manageable parts ("piles" of similar small multiples). These piles can be interactively compared to reveal temporal states such as stable periods or transition between these periods. A different approach has been taken by Senk at al. in their tool VIOLA [SCH+18]. In this application, the state of a simulated neuronal-network activity in a 2D neuronal layer can be traced over time. A 2D heatmap is visualized over time in a 3D volumetric rendering (i.e., the third dimension is time). The user can manually select individual time points to get detailed information about how many neurons fired and at what rate in other views. Another, although not neuroscience-related, approach to target this problem has been proposed by van den Elzen et al. [vdEHBvW16]. Here, snapshots are projected to a low-dimensional space as points via dimensionality reduction methods (Figure 2.9B). This "time map" enables the identification of stable or recurring states that appear as local clusters, which might be relevant for finding temporal patterns.

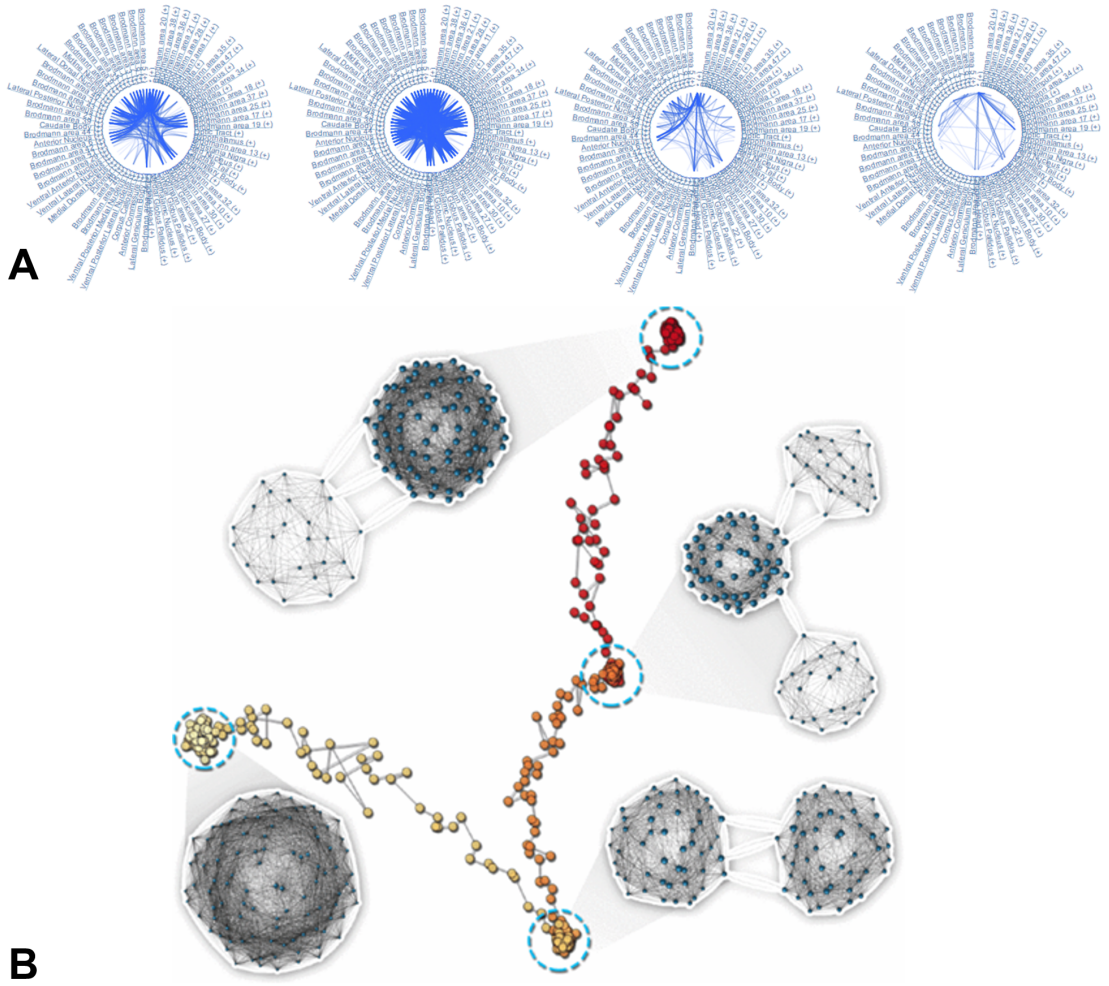


Figure 2.9: Examples for network comparison: (A) Small multiples of four functional networks in a circular layout (brain regions) ([dRKY+18], Figure 3). (B) 125 snapshots of a dynamic network (i.e., 125 states of a network at different time points) mapped to a two dimensional space via PCA ([vdEHBvW16], Figure 8). Stable states (clusters) are highlighted with blue circles (gray graphs show a representative snapshot).

2.2.3 Exploration of Spatial Big Brain Data

Neuroscience studies that use a combination of imaging, anatomical, and connectivity data often require extensive analytical workflows involving manual data aggregation and statistical analysis to find patterns in big brain data [FP11, JFD14, RA15, RYWB15, WVRG+16, FTP11, FF16]. The term "big" refers to the amount (vast image collections) and/or size (high resolution image/network data) of the data which is too complex to analyze with traditional methods (note that there is no general consensus about the term "big data", the Oxford Dictionary defines it as "extremely large data set").

Visual analytics tools may support neuroscientists in this task beyond non-interactive visualizations that do not interactively link data, such as BrainBrowser[SKR+15] or VisBrain [CVO+19].

Exploring a large collection of imaging data in combination with anatomical data can provide spatial context and orientation. BrainScope [MvdGvdM+15, MHLR16, HvM+17] uses 2D slices of anatomical parcellations derived from the Allen Human Brain Atlas [HLGB+12] to visualize gene expression on the web. For this purpose, it utilizes a genome-wide collection of region-level spatial gene expression data [HLGB+12]. To make the entire collection visually comprehensible, they use t-SNE (t-distributed stochastic neighbor embedding) to map the genes to a two dimensional space based on their correlation (Figure 2.10A), presented in a scatter plot. This mapping is linked with 2D slice views, so selections of genes are visualized by their average regional expression. Vice versa, a selection in the slice views leads to a color coding of the scatter plot to highlight genes with high/low gene expression in the respective brain regions. A similar approach is used in INVIZIAN [BJV12] for a collection of 900 anatomical MRI images of humans. In this paper, the images are mapped via MDS (multidimensional scaling) to a low-dimensional space based on a distance metric concerning cortical patterns in these images. In contrast to BrainScope, individual images are not rendered as dots in a scatter plot but as cortical surfaces in a 3D space (i.e., a 3D "cloud" of brain meshes). Detailed information, such as slice views of the respective images, can be retrieved via selecting individual brains. The cloud can be color coded via textual queries or a parallel coordinate system to visualize additional features like sex, age, test group, association etc.

High-throughput electron microscopy imaging platforms can create high resolution imaging data with cellular resolution [ZLP+18]. Tracing neurons in this data can be used to create structural connectivity on a micro-circuit level that can ultimately lead to "wiring diagrams" of the brain. This requires tools to accurately segment neurons in large volumetric imaging data in the range of terabytes called EM stacks, a collection of 2D slices of electron microscopy images. One of them is CATMAID, introduced by Saalfeld et al. [SCHT09, SMGL+16]. CATMAID's primary purpose is manual neuron annotation in large arbitrary EM stacks from different species. For this, skeleton traces of neurons are drawn across 2D slice views. These can be further combined to networks. These networks can be rendered in a separate view as node-link diagram with flexible layouting algorithms. Selections are linked between several parallel views, including 3D representation of the neurons, 2D slices, graphs, statistics, etc., so that the user does not lose the context to the spatial representation. ConnectomeExplorer [BAAK+13] took a different approach. Instead of focusing on the annotation and segmentation of EM stacks, Beyer et al. presented a tool for the analysis of the data that had been produced in the this process. To manage the collection of EM stacks, segmentations, annotations, connectivity data, and meta data to answer domain specific questions, they proposed query-guided interactions. This can be done in a visual query builder - a user interface element - to ask for computed, manually labeled, or topological attributes of

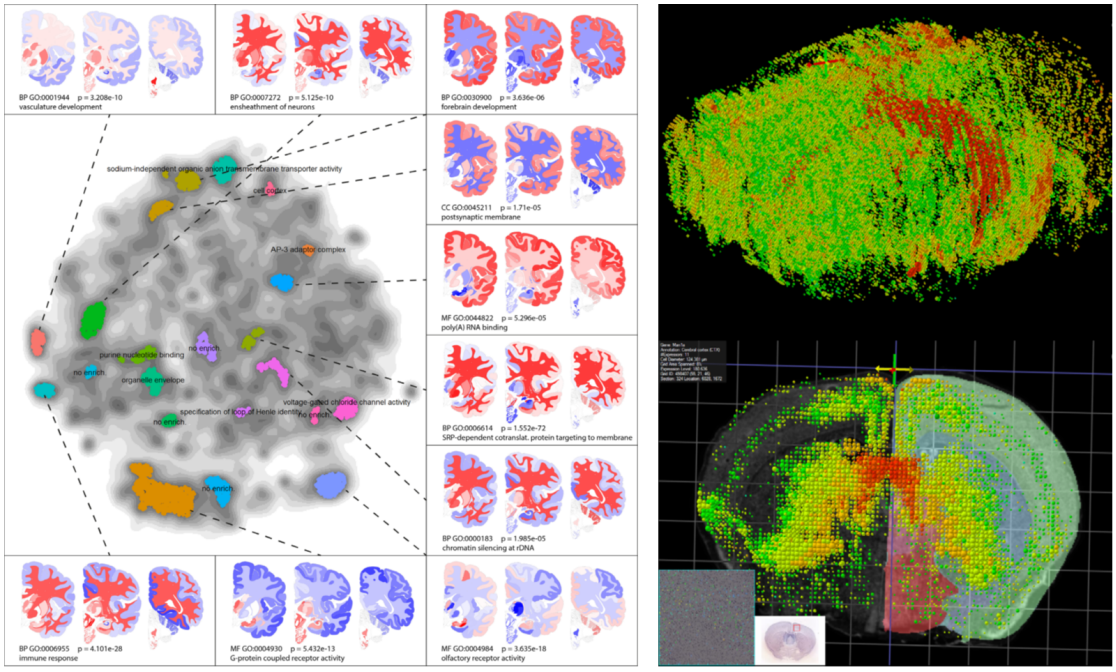


Figure 2.10: Visualization of spatial gene expression data: (A) Mapping of 3000 genes to a two dimensional space via t-SNE of their gene expression ([HvM+17], Figure 3). Expression of representative genes of several clusters are shown in 2D slice views of the human brain. (B) Spatial gene expression of the gene Man 1a via glyphs, overlaid with anatomical context of the mouse brain ([LNT+08], Figure 4a and c).

the data, or by the data object's distance to a region of interest. Query results can then be visually explored in linked views, comprising a 3D volume/mesh rendering, a 2D slice view, connectivity graphs, a tree-view showing the hierarchical structure of segmentations, and several statistical views (histograms, scatterplots etc). Since the 3D rendering of the segementated neurons causes cluttering and obstruction (Figure 2.11, left), Al-Awami et al. integrated NeuroLines [AABS+14] into the ConnectomeExplorer. NeuroLines transforms these segmentations to a 2D abstract visualization that resembles a "subway map" (Figure 2.11, right). The map preserves the underlying anatomical tree structure, as well as its distances. Hence, it provides a spatial representation of the neurons without the complexity of a 3D visualization.

Another query-based approach has been introduced by Bruckner et al. in BrainGazer [BSG+09]. Here, they used visual queries to explore large databases of transgenic fruit flies that consist of volumetric imaging data acquired through confocal microscopy as well as segmentations of neuronal structures (Figure 2.12). Visual queries enable to retrieve contextual information by selecting ROIs (regions of interests) directly in 2D/3D rendering views. This information can be either semantic (information from the database)

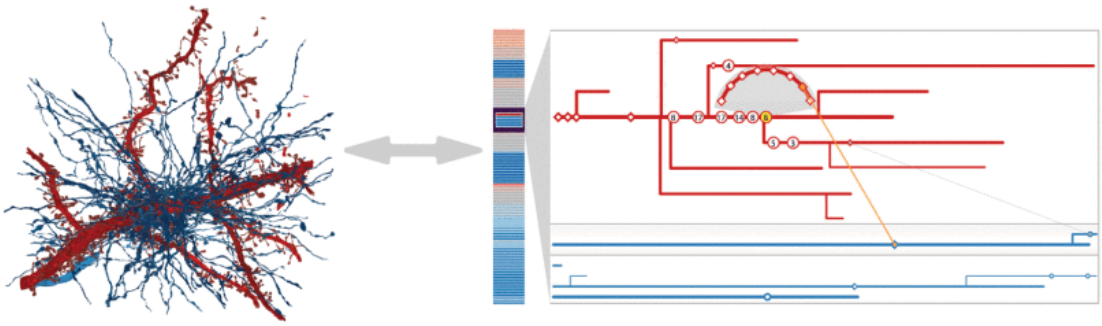


Figure 2.11: 3D segmented dendrite (red) and a connected axon(blue) (left) mapped to an abstract representation similar to a "subway map" (right) with NeuroLines ([AABS+14], Figure 1).

or spatial (objects with a spatial relationship). ROIs can be either neuronal structures or arbitrarily selected with a freehand drawing interface. Hence, queries are not limited to anatomical structures. Furthermore, neuronal structures can be mapped to interactive wiring diagrams with NeuroMap [Sor13], an extension to BrainGazer previously described in Section 2.2.1. These diagrams provide a 2D overview of precomputed pairwise overlaps of neuronal structures to show synaptic connectivity, i.e., if two neurons are structurally connected). To detect these overlaps, a problem that leads to visual clutter if more than two neuronal structures are involved, Swoboda et al. [SMB+17] extended BrainGazer with an interactive overlap detection for multiple neuronal structures. A glyph-based abstraction of a real-time GPU-based computed overlap is rendered onto a 3D anatomical representation of the brain. Glyphs indicate which structures overlap and to what extent. With this information, neuroscientists are able to gradually adapt the knowledge about neuronal wiring. By now BrainGazer has been further developed into Brain* [bra19], a web-based software framework to manage, access, and visualize large collections of confocal microscopy images. It provides the basic features of the original BrainGazer application, such as visual queries, volume rendering, and mesh rendering, without the need for a local installation. Hence it can be used to provide access to public resources [lar19, fru19, zeb19].

Neuron Navigator is a tool following an approach similar to BrainGazer and has been introduced by Line et al. [LTW+11]. Neuron Navigator can query connectivity of and between ROIs in the fruit fly's brain space. These ROIs can be either neuronal structures or arbitrary cuboid boxes, selected in a 3D rendering view. Neuronal structures that reflect connections from, to, or between ROIs can be retrieved via queries that access a 3D neuron image database and matching the region of interest with annotated locations of neuron terminals.

Another tool for querying connectivity data is BrainExplorer [LNT+08, FLN+15]. This tool enables the retrieval of incoming/outgoing structural connectivity from the Allen Mouse Brain Connectivity Atlas [OHN+14] at brain region-level. BrainExplorer utilizes

a hierarchically organized brain ontology [LHA+07] to visualize brain anatomy in 2D and 3D. Brain regions can be selected to retrieve tubular trajectories of outgoing structural connections (Figure 2.13) that have been pre-computed from imaging data of the Allen Mouse Brain Connectivity Atlas [OHN+14]. Furthermore, BrainExplorer provides a glyph-based 3D visualization of gene expressions mapped to a standard brain (Figure 2.10B), where the color and size of the glyphs indicate the gene expression level. The tool enables the execution of explicit gene queries, and searching for specified genes of interest based on specific anatomical regions of interest. For identified genes of interest, a correlation query can be executed, returning genes with similar gene expression. Since anatomical brain regions can be queried for either structural connectivity or genes, a combination thereof represents a linking of imaging data with connectivity data on a region-level.

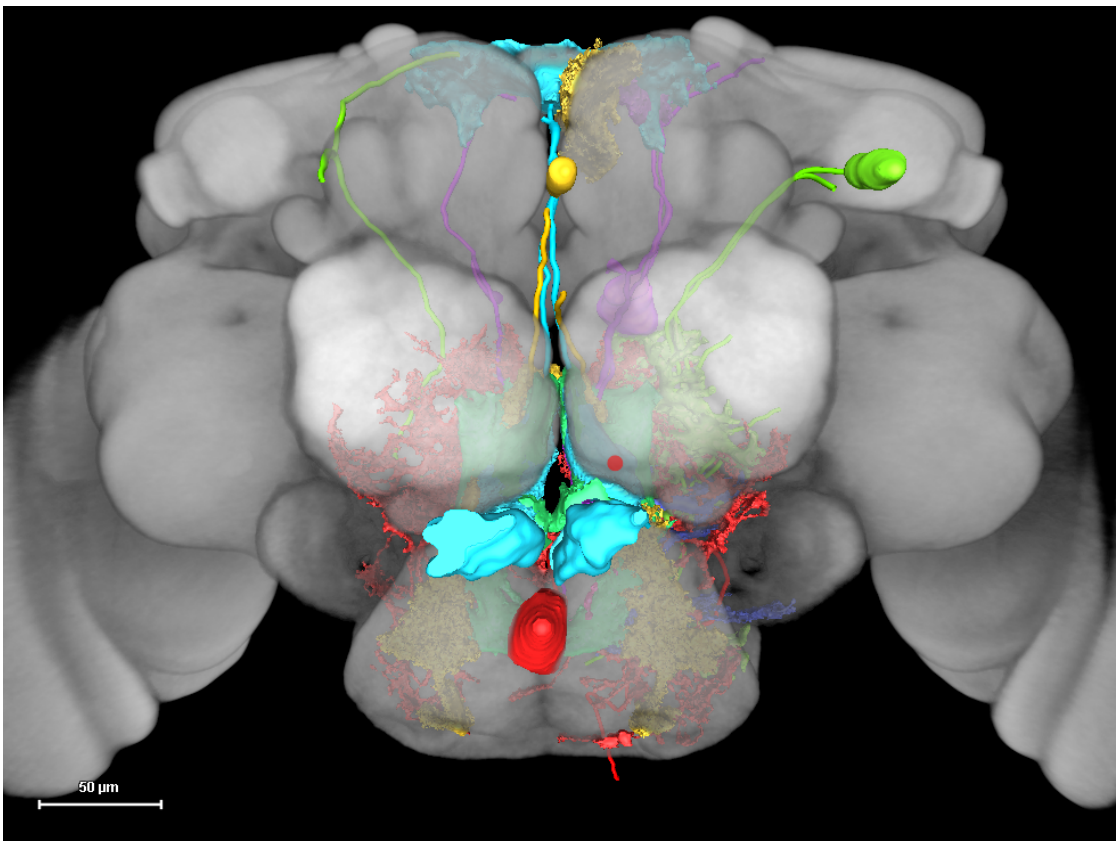


Figure 2.12: 3D rendering of segmented neuronal structures in the fruit fly's brain ([SMB+17], Figure 2).

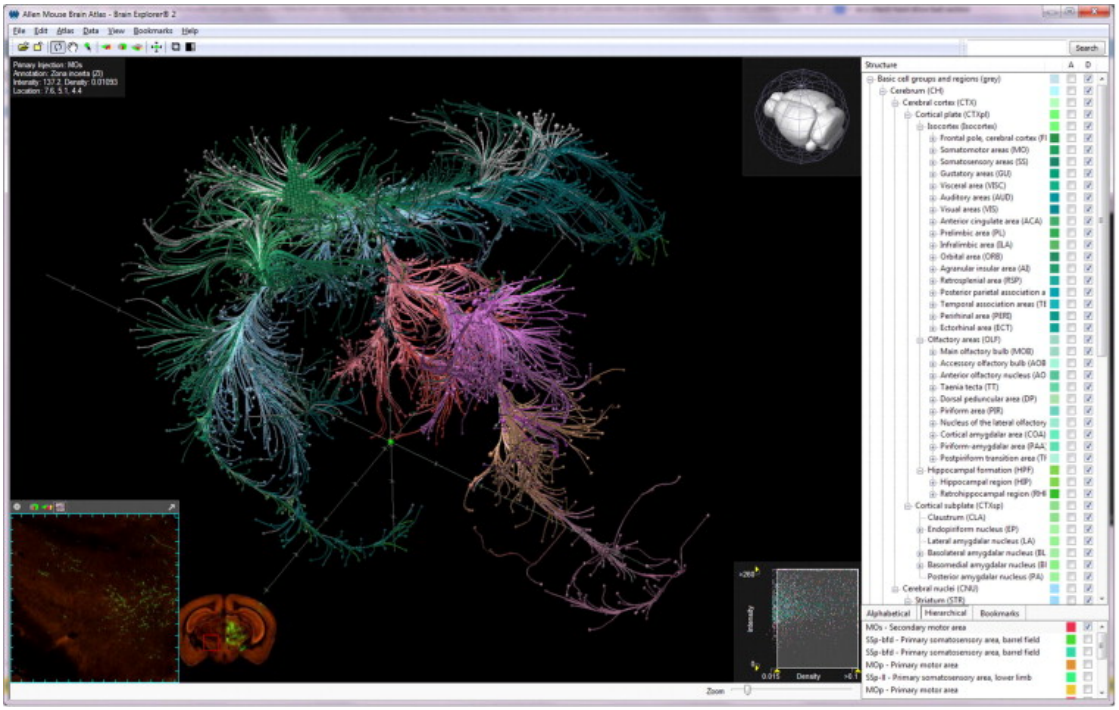


Figure 2.13: 3D visualization of neuronal structures and connections. Connections emerging from a specific brain area (secondary motor area) visualized as tubular trajectories. Color codes indicate the brain regions passed by the connections ([FLN+15], Figure 1).

CHAPTER 3

Detailed Contributions

3.1 Paper A: Predicting Functional Neuroanatomical Maps

Genetically weighted connectivity analysis (GWCA): Most established approaches that map genetic information to brain connectivity, relate gene co-expression correlation of functionally grouped genes with structural connectivity [FP11, JFD14, RA15, RYWB15, WVRG$^+$16, FTP11, FF16]. A correlative analysis primarily reflects transcriptomic similarities, i.e., if similar genes are expressed, but does not represent functional synergies of multiple genes. Hence, we sought to develop a novel computational method to explore these synergies and how they affect connectivity. The genetically weighted connectivity analysis (GWCA) (Figure 3.1) is based on the hypothesis that the functional synergies of a set of genes are reflected in their cumulative impact on brain networks such as global structural connectivity or functional resting state networks. Therefore, we determine the cumulative expression of the gene sets (Figure 3.1, 1 and 2), represented by the voxel-level mean gene expression of the genes, and use it as weighting for the global connectivity (Figure 3.1, 3). The impact is measured by computing voxel-level higher-order features of this network, i.e., node-based network measures such as node strength, hubs, authorities, closeness, betweenness, and eigencentrality (Figure 3.1, 4 and 5). A combination of these features eventually represents the functional neuroanatomical map. Statistical comparisons to random gene sets determine how significantly these maps are associated with brain function/behavior. We have shown the validity of this approach by comparing the predicted functional maps with functional neuroanatomical annotations from literature and fMRI data.

High-throughput analysis: Our methodology can be applied to model organisms for which gene expression maps, connectomes, and genetic information is available. Functional maps from a single gene set can be computed in 1-2 hours on a machine with 30 CPU cores. There is no need to tune parameters for different genes sets, so processing

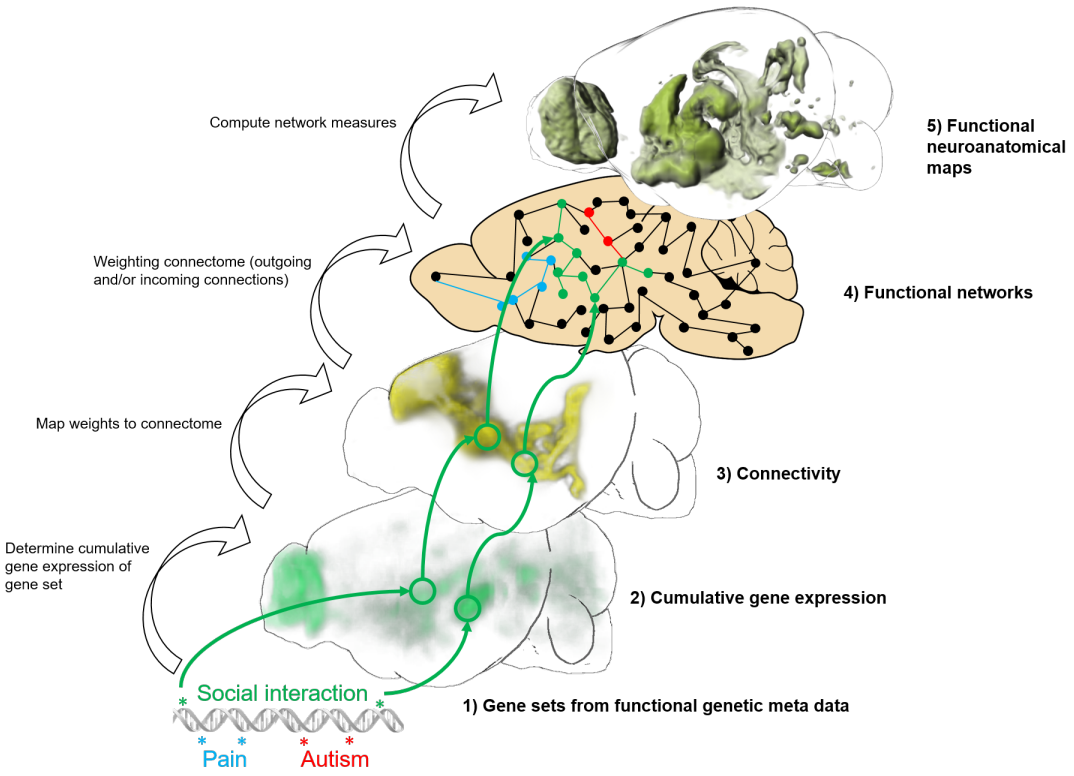


Figure 3.1: Concept of GWCA: From genetic meta data to functional neuroanatomical maps

multiple gene sets can be performed as a batch. An example for functional maps generated from nine different gene sets can be seen in Figure 3.2. When applied to large collections of gene sets from different sources, such as genome wide association studies (GWAS), quantitative trait loci (QTL) analyses, or neurogenetic databases, our approach can rapidly screen for neural circuitry underlying specific brain functions, behavioral traits or psychiatric diseases, and therefore refine the known functional organization of the brain.

A genetic algorithm to refine gene sets for functional synergies: Gene sets from multigenic meta data may contain genes that do not show synergy with others, do not show gene expression in the brain, or are ubiquitously expressed. This can lead to an unspecific synergy, which shows only low contrast in the predicted maps, and consequently low significance. Therefore, these genes need to be removed from the sets. We solved this optimization problem by employing a genetic algorithm that maximizes the variance of the cumulative gene expression, so the sets contain only genes that have high expression at the same brain regions. We showed empirically that a higher variance led to a higher contrast in the predicted functional maps that were - apart from the enhanced contrast - similar the original gene sets.

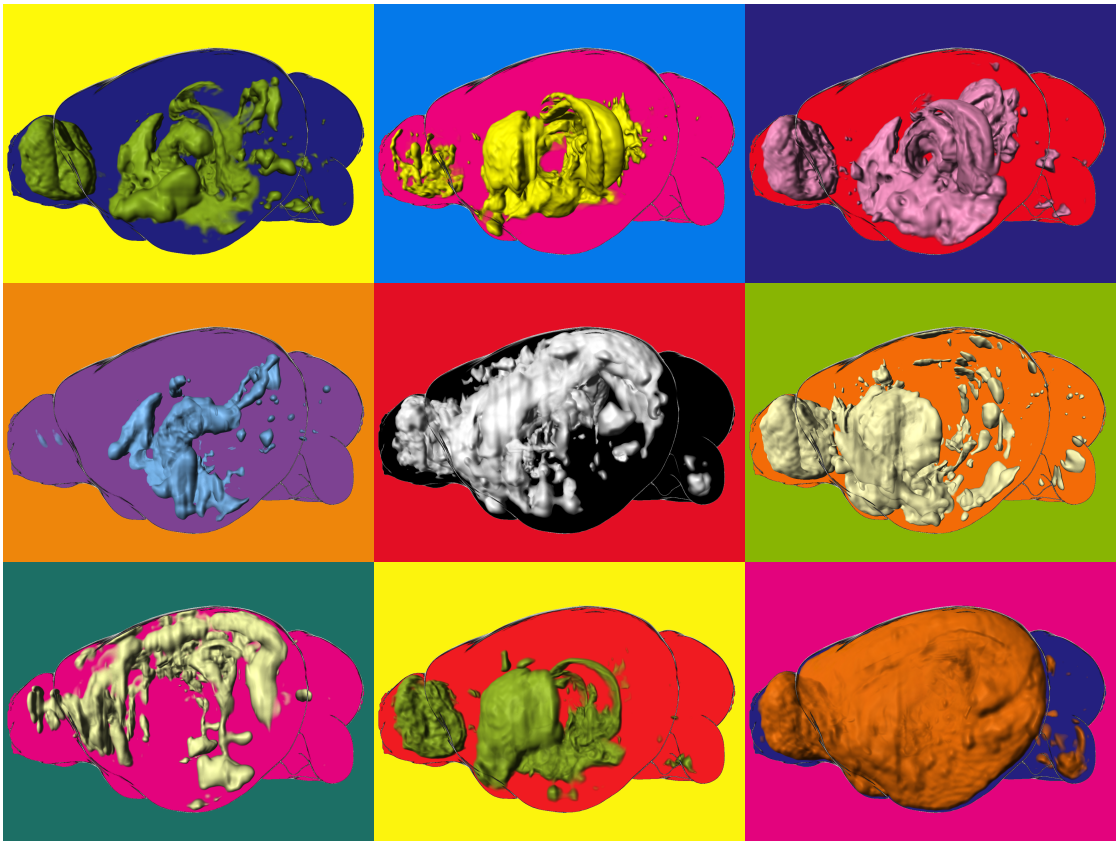


Figure 3.2: Functional maps generated with GWCA from functional gene sets associated with the central amygdala circuitry, dopaminergic signaling, feeding, hypothalamic circuitry, fear memory consolidation, panic disorder, learning in a stressful context, social bonding, and synaptic plasticity (left to right, top to bottom). The colors emphasize the versatile nature of the gene sets and have no specific meaning.

Potential in emulating functional fMRI data: We compared functional maps of pain-related gene sets to fMRI recorded during pain perception in a mouse. Domain experts confirmed a high similarity between these two (Figure 3.3). Since this represents only one sample, one cannot draw a general conclusion. Since several pain-related gene sets correlated with this one fMRI image, one can at least see the potential to simulate fMRI in silico, in cases when no functional brain data is available.

Visualization of brain networks with spatial, neuroanatomical context: The predicted functional neuroanatomical maps show the effect of multigenic behavior/functions of brain connectivity on a voxel-level scale (as already shown in Figure 3.1). To identify which candidate circuits may be involved in the function/behavior, these maps need to be explored in the context of the connectome. Hence, we visualized structural connectivity between brain regions highlighted in the maps. To create a visualization that is suitable

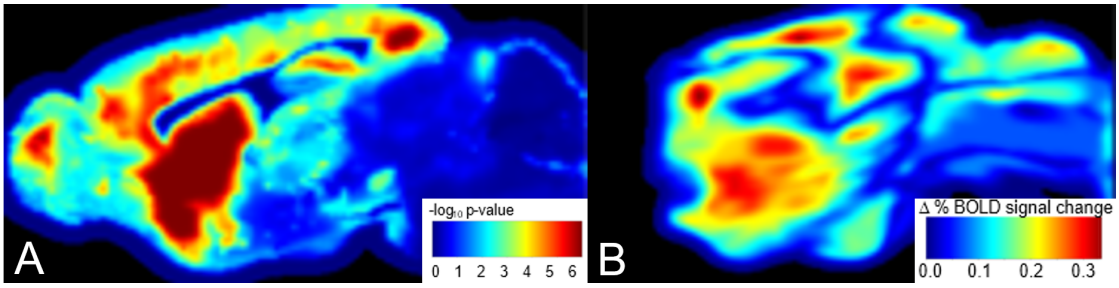


Figure 3.3: Comparison of functional maps and fMRI of a mouse brain. (A) A slice of a functional map related to pain. Color indicates how significant a voxel is associated to brain function/behavior. (B) A slice of an fMRI image related during pain perception. Color indicates a BOLD (blood oxygen-level dependent) signal change between control and pain. The olfactory bulb (front part of the mouse brain) was not available, since it has not been imaged during fMRI.

for figures in publications, we performed several design iterations with neuroscientists. This progress was not in the scope of the *NeuroImage* journal, so we did not report it in Paper A. Nevertheless, we want to comment on it in this section because it might be of interest for the visualization community.

At first, we performed informal research with neuroscientists, how neural circuits are visualized in the literature. Figure 1 in Russo et al. [RN13] represents a common abstraction of neural circuitry in neuroscience (Figure 3.4A), which our collaborating neuroscientists felt intuitively familiar with. To get brain regions/connections similar to the figure, it was necessary to cluster the functional map. A visualization of a voxel-level network (with billions of edges) or on the anatomical brain region-level (still hundreds of edges), would create too much clutter in a 2D figure. Therefore, we clustered the voxels of the functional map (i.e., voxels with significant association with function/behavior) by structural connectivity, so that the connections of the voxels between clusters bundle together.

Our first approach was to use a sagittal projection of the mouse brain where colors indicate different clusters, and render the connectivity (sum of all connections) between them as directed arrows (Figure 3.4B). Short names of the main anatomical regions for each connection are listed in the top-left legend. Although it was well received by our domain experts, we rejected it after several discussions due to cluster and edge obstructions. As a solution to this problem, we split the clusters into separate sagittal projections (Figure 3.5A), and visualized the connectivity in-between. The brain regions of the clusters are indicated as circles, where their sections show relative region size to represent a direct anatomical context. The sections were labeled by their short name and color, each color is associated with a major brain region according to the Allen Brain Atlas. Another advantage compared to the first concept is that one can visualize additional voxel-level information, e.g., how significantly a voxel is associated with function/behavior. We

discussed this visualization with several neuroscientists, who found it too "overloaded" for a figure in a paper, with the voxel/circle coloring especially generating confusion. Therefore we removed the circles and replaced them with labels (Figure 3.5B).

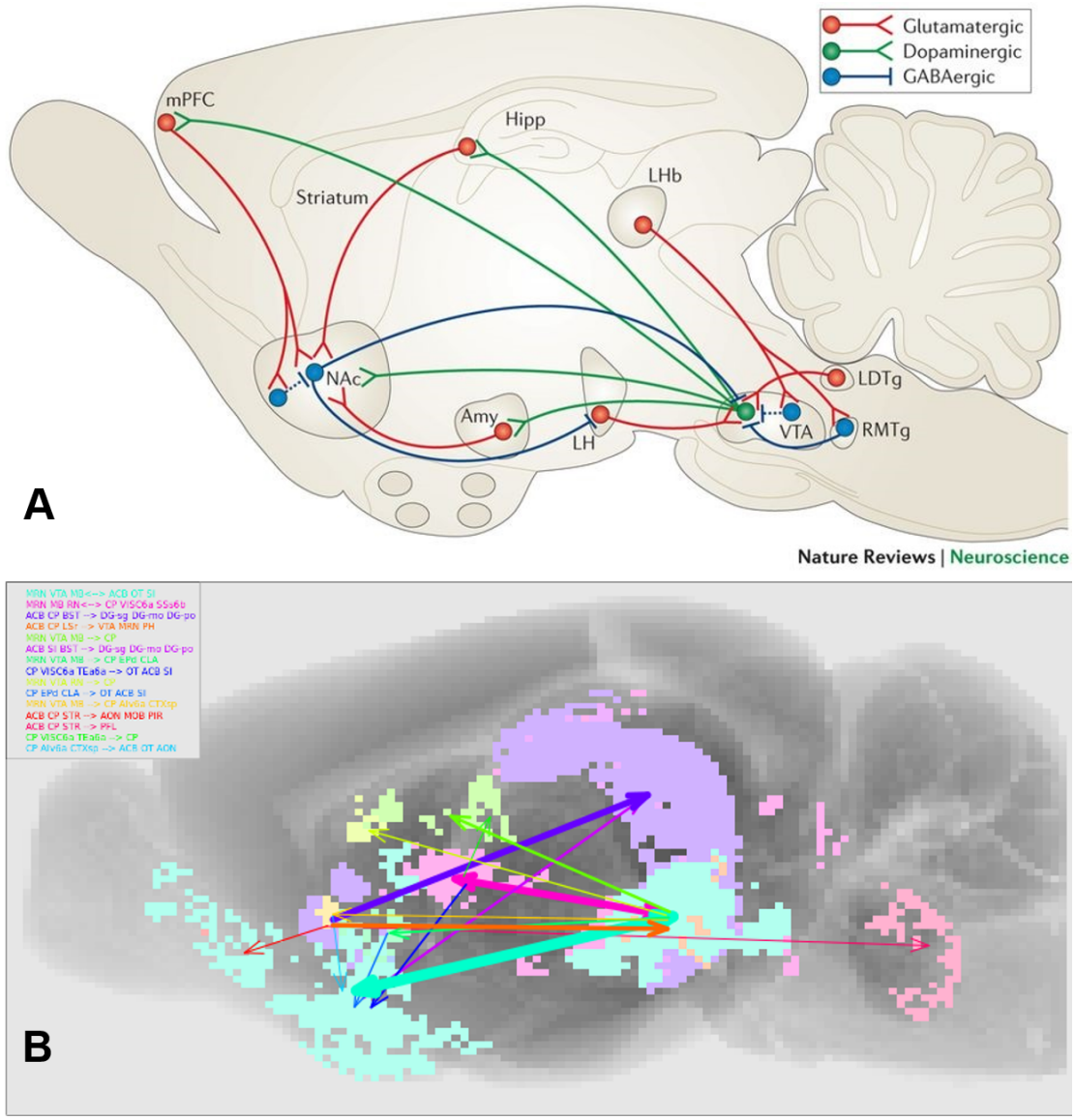


Figure 3.4: Concepts for a visualization of brain networks within neuroanatomical context. (A) Schematic of brain reward circuitry in a mouse brain by Russo et al., Figure 1 [RN13]. (B) Sagittal projection of the mouse brain. Clusters of the functional maps are colored differently, the short names of the clusters' major anatomical regions are listed in the legend.

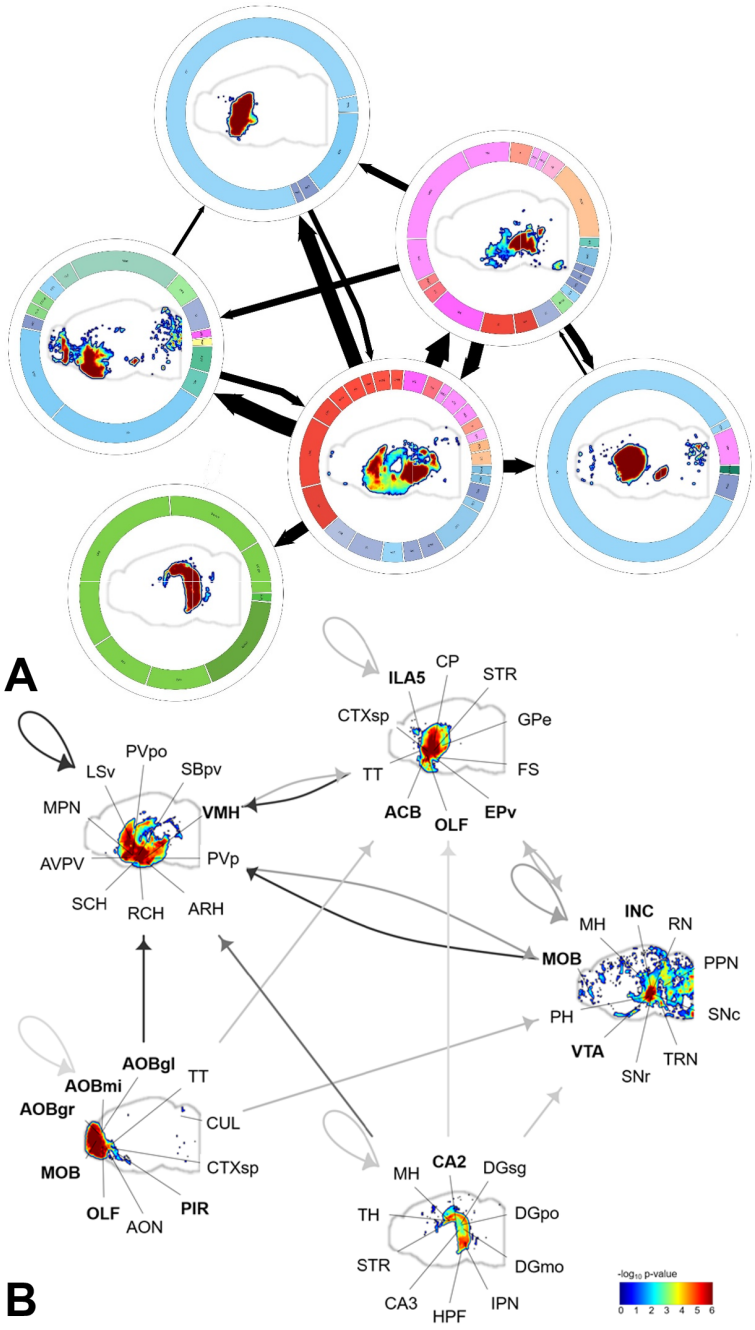


Figure 3.5: Different design stages for a visualization of brain networks with spatial, neuroanatomical context: (A) Sagittal projections of six clusters of functional maps. Voxel color represents how significantly the voxel is associated with a specific behavior/function. Colored circles show the relative number of voxels in the respective brain region. Arrows indicate the structural connectivity in-between. (B) Sagittal projections of five clusters of functional maps, similarly to A. Labels show the major brain regions of each cluster.

3.2 Paper B: A Data Structure for Big Brain Networks

Real-time querying of large-scale spatial networks: Exploring the connectivity of brain regions or volumes of interest - a set of voxels representing a user-selected part of the brain - on a voxel-level involves querying large graphs for connections where either their sources (origin) or targets (destination) are spatially close to each other. In recent years, a multitude of graph engines have been published [KBG12, HLP+13, RMZ13, CDW+16, CDW+16, CDW+16, ZMB+15] that are able to handle graphs that do not fit into main memory. Despite their universal applicability, they are not tailored to take advantage of spatial organization to enable real-time queries. Hence, we developed a novel data structure that stores the outgoing and incoming connections of neighboring voxels close to each other on the hard-drive, so that sequential read-speed can be achieved. Hard-drives and SSDs are optimized for sequential read-speed, so this represents the fastest way to read data. Therefore, we arranged the connectivity data along a Hilbert curve (Figure 3.6). Combined with a file format that exploits the sparsity of the data, our data structure is able to outperform state-of-the art graph engines at *Aggregation Queries*, which will be further discussed in the next paragraph.

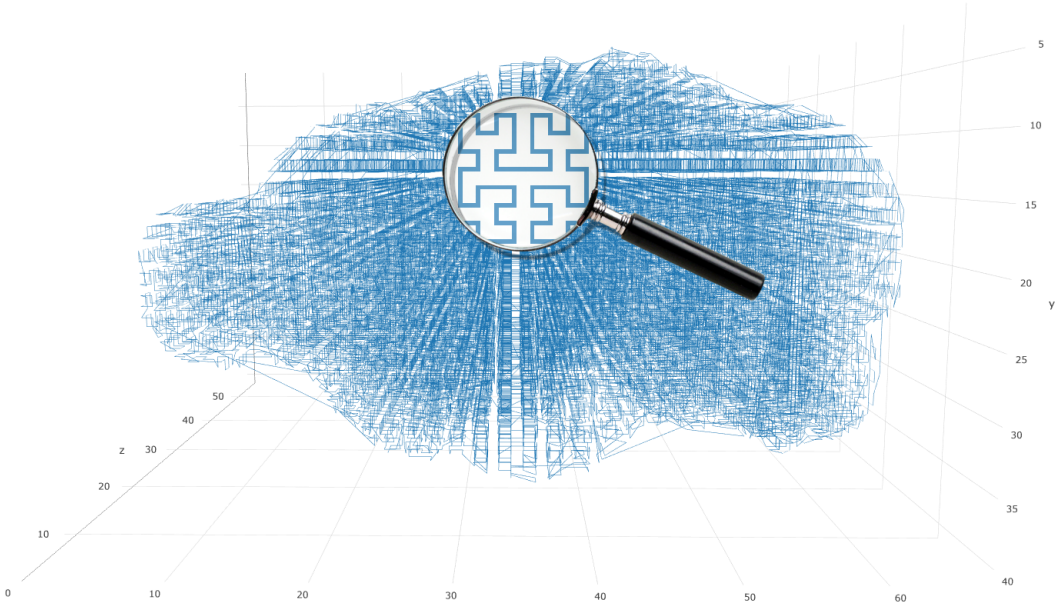


Figure 3.6: A Hilbert curve through a mouse brain volume. Along this curve, locality is preserved.

Instant retrieval of connectivity across scales: Bridging the gap between different resolutions and scales requires a mapping of connectivity from one level to another one, e.g., from voxel-level to region-level connectivity. For this purpose, our data structure provides a hierarchical mapping to a common reference space. It utilizes hierarchical brain parcellations, where the lowest level represents voxels of the reference space and

higher levels represent brain regions (Figure 3.7). Hence, it is known, which voxels belong to which region, so region-level connectivity can be interpolated to voxel-level, and vice versa, voxel-level connectivity can be cumulated to region-level connectivity. Since this aggregation may involve voxel-level connectivity with potentially gigabytes of data, we envisioned *Aggregation Queries*. In these queries, where the underlying concept was described in Figure 1.3, connectivity from, to, or between arbitrary volumes of interests (user-selected or brain regions from the hierarchy) are aggregated on the voxel-level. For local brain queries - a sub-brain region size that covers less than one percent of the brain - the query can be executed in less than one second on a 100 gigabyte connectivity matrix integrated in the data structure.

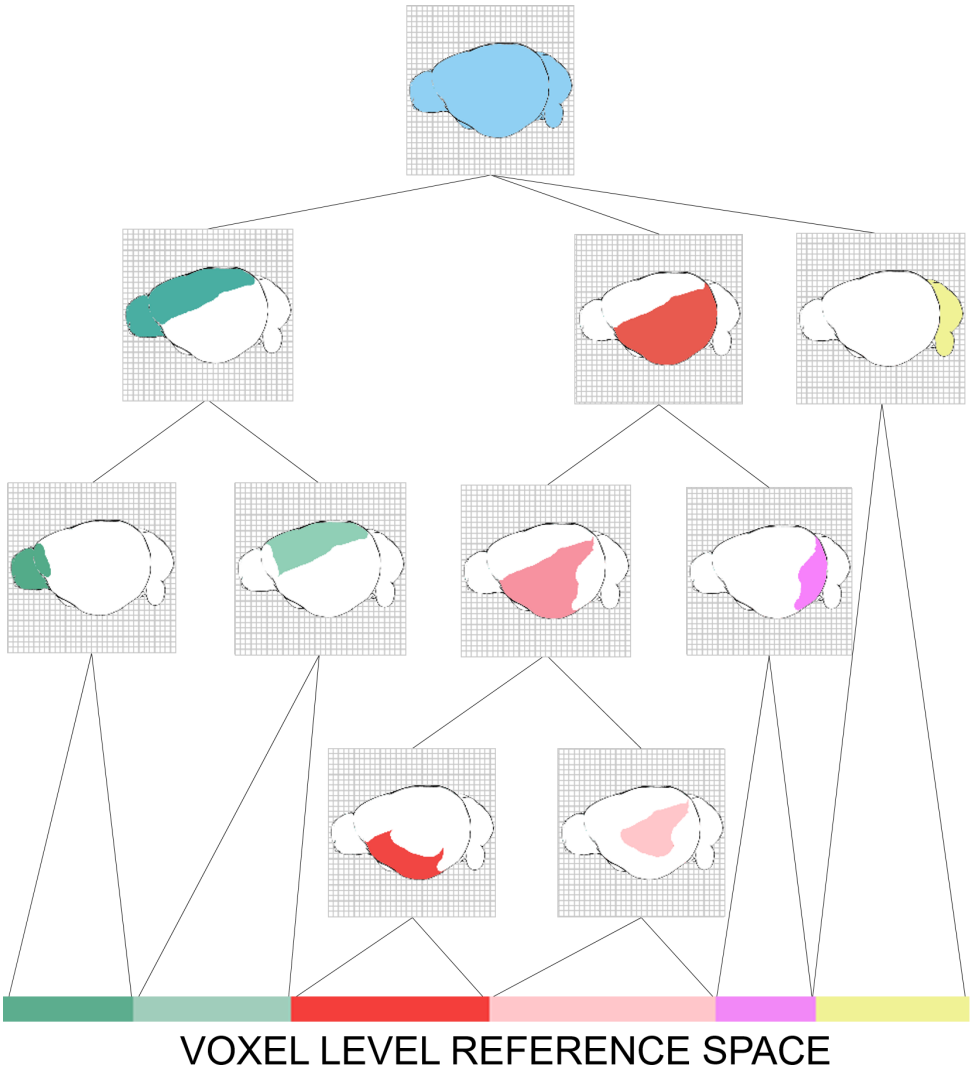


Figure 3.7: Schema of a hierarchical mapping: The lowest level represents a voxel-level reference space, while higher levels comprise brain regions.

Scalability: For volumes of interest covering more than one percent of the brain, we employed query caches. These caches contain pre-executed queries of brain regions or blocks (neighboring voxels on the reference space). Whenever a query is executed on the data structure, the caches will be checked first if parts of the volume of interest are already computed. If this is the case, only the remaining voxels need to be read from the data structure. Although this creates a storage space overhead, it enables queries of arbitrary size to be executed in an instant. Furthermore, these caches can cover different scales to form an equivalent of image pyramids - one cache with $\frac{1}{2}$ of the resolution of the original connectivity matrix (8 voxel blocks), another one with $\frac{1}{4}$ of the original connectivity matrix (64 voxel blocks) and so on. Hence, queries can scale on even larger data.

Enabling hierarchy based-navigation schemes: Retrieving brain networks on a global level, i.e., brain-wide region-to-region connectivity, would require multiple *Aggregation Queries* for each brain region. Since even one *Aggregation Query* can take a second, it is not feasible to compute global connectivity at runtime. Instead, we store the global region-level connectivity across all regions of the hierarchical brain parcellation (Figure 3.7) in a graph-database. This allows visual analytics workflows to explore global region-level connectivity on different hierarchical levels (Figure 3.8). Consequently, neuroscientists can interact with a global network of brain regions that fit their research interest (e.g., cortical regions rather than the cerebellum). We demonstrate with a prototypical web-component, that this can be even used for inter-species comparison of networks, with network similarities identified by comparing the networks iteratively on different anatomical levels in parallel between the mouse and the human brain.

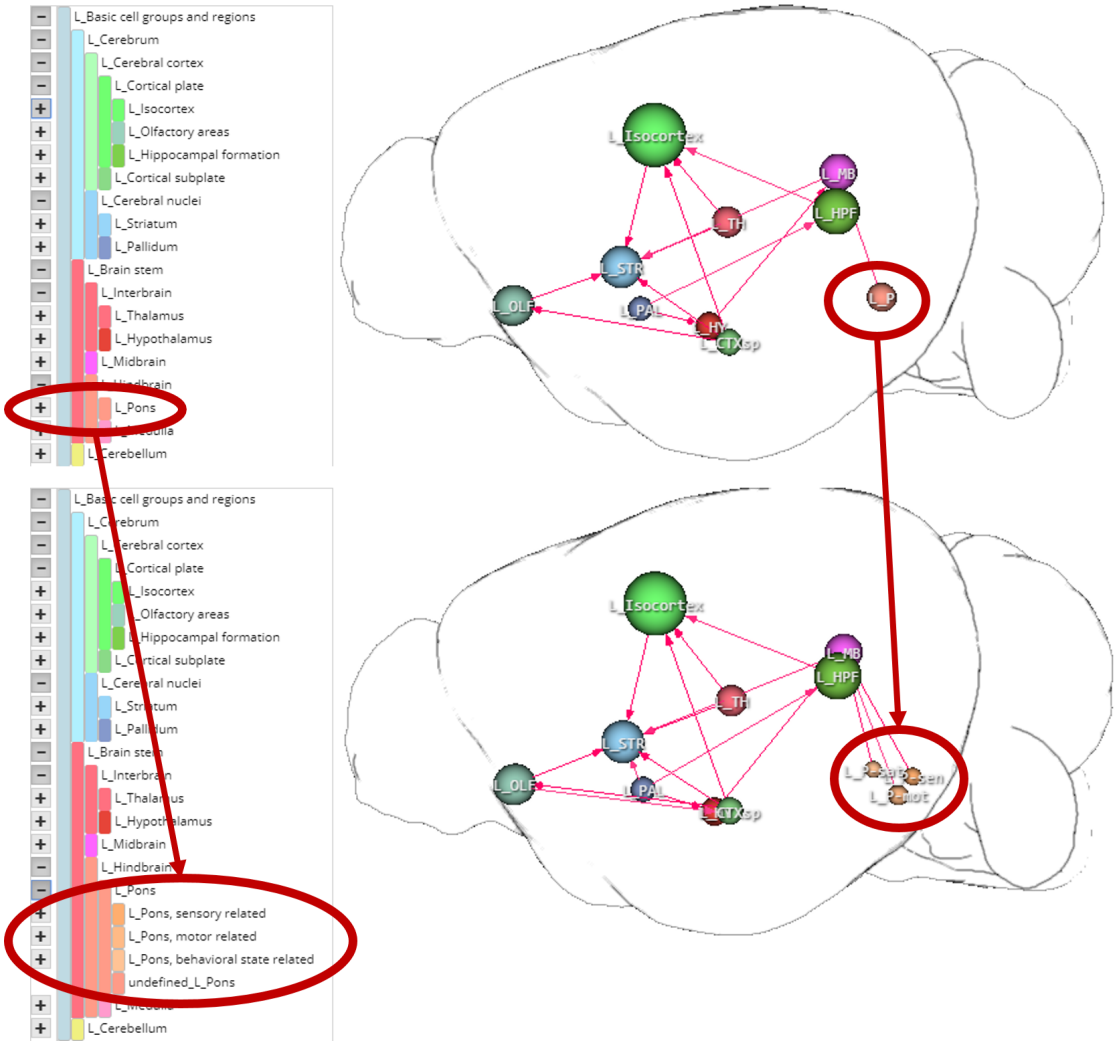


Figure 3.8: Network visualization on two different hierarchical levels: The top panel shows the region L_Pons, while the bottom panel shows the sub-regions of L_Pons.

3.3 Paper C and D: BrainTrawler

Paper D represents an extension of Paper C. Hence, the contributions of Paper C are fully included in Paper D. Contributions that are solely part of Paper D are stated explicitly in the text.

Relating spatial data to connectivity: Brain regions that are connected to expression sites of a gene (Figure 1.1) may also contribute to the gene's function or be a second-order effect thereof. Linking spatial gene expression data with structural, functional, or other kinds of connectivity is therefore a first step in revealing the genetic foundation of neuronal circuits. In contrast to Paper A, we targeted this problem with visual analytics methods. For this, we envisioned an interactive selection of gene expression sites that are relevant for neuroscientists (e.g., within a certain brain region of interest) and employed *Aggregation Queries* to identify the outgoing connections (*Target Connectivity Query*) and the incoming connections (*Source Connectivity Query*) to other parts of the brain - so-called target/sources regions of the expression sites. Although the principle of this approach was demonstrated in Paper B, it lacked the inclusion of an anatomical context in the selection process and the connectivity visualization. This would allow neuroscientists to perform these queries in a more accurate manner in relation to the brain anatomy. Therefore, we used 2D slice views of the brain to visualize an anatomical atlas with brain regions as outline, overlaid with gene expression data (Figure 3.9A). In these views, selections of voxels with high gene expression can be performed either manually with brush-drawings, or by thresholding within brain regions (Figure 3.9B). The selections act as volume of interest for an *Aggregation Query*, where the resulting voxel-level connectivity reveals the target/source regions in 2D and 3D views (Figure 3.9C and D). Quantitative information is given in *Connectivity Profiles*, a bar chart that shows the mean connectivity within the brain regions corresponding to the slices views (Figure 3.9E).

Visualization of connectivity on different scales: Mapping voxel-level connectivity to different anatomical scales provides a different level of abstraction. Since different brain regions - from small nuclei to large cortical layers - might be relevant for neuroscientists, a region-level representation depends on the user-selected level of detail. We targeted this problem with a hierarchy-based navigation-scheme, as presented in Paper B. This requires pre-computed region-level connectivity on different hierarchical levels, which is not feasible for arbitrary user-defined volumes of interest (i.e., a volume of interest may span several anatomical regions, which cannot be split anymore after an *Aggregation Query*). To solve this problem, we developed *Split Aggregation Queries* in Paper D, which divide a volume of interest into anatomical sub-regions and aggregate the connectivity separately. The concept of a *Split Aggregation Query* can be seen in Figure 3.10. At first, the volume of interest is split to the lowest level of the hierarchical brain parcellation. A *Split Aggregation Query* returns voxel-level connectivity for every region within the volume of interest. This can be combined to the voxel-level connectivity of the whole volume of interest (equal to an *Aggregation Query*) or region-level connectivity. This saves the overhead created by executing several *Aggregation Queries* sequentially and

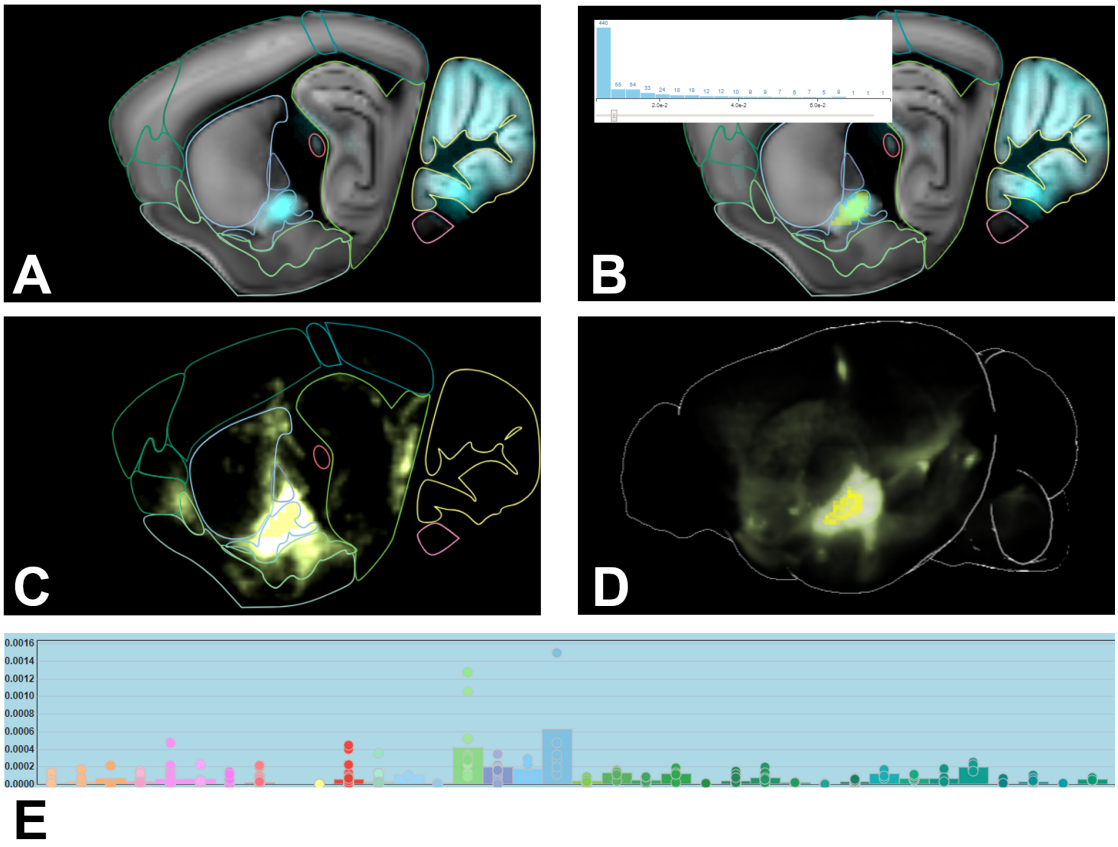


Figure 3.9: Relating gene expression data to structural connectivity: (A) Visualization of gene expression of the gene PKC-delta (cyan) in a sagittal slice of the mouse brain (anatomical brain regions are outlined). (B) Selection of voxels with high gene expression (yellow) within the Striatum-like amygdala nuclei region (blue outline). (C) Outgoing structural connectivity of the selection (light green). (D) Outgoing structural connectivity of the selection in 3D. (E) *Connectivity Profile* shows the mean connectivity of brain regions as bar chart. Circles represent the connectivity for anatomical sub-regions.

achieves a performance similar to a single *Aggregation Query*. Hence, it is possible to visualize voxel-level connectivity on different levels of a hierarchical brain parcellation (Figure 3.11).

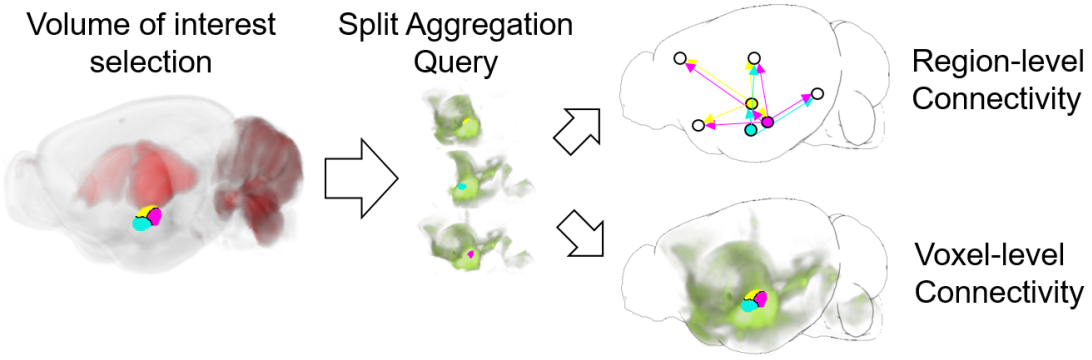


Figure 3.10: Concept of a *Split Aggregation Query*

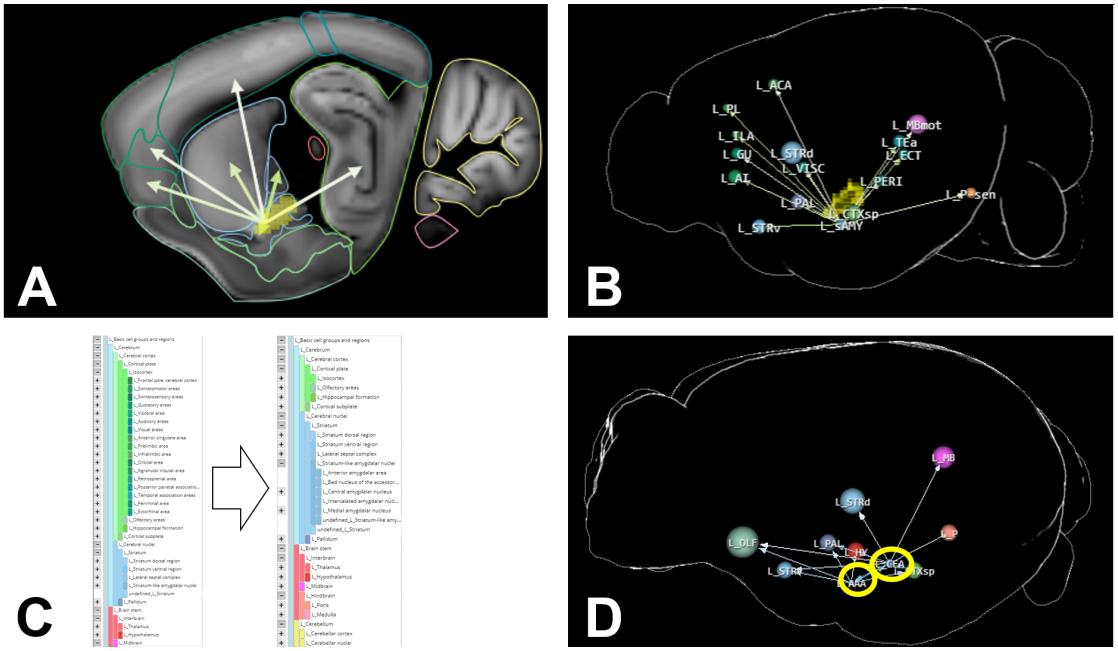


Figure 3.11: Voxel-level connectivity of Figure 3.9 on the region-level: (A) Outgoing structural connectivity of the selection on region-level. (B) Outgoing structural connectivity of the selection on region-level, represented as 3D node-link diagram. The center of the spheres indicate the regions center-of-mass, the size encodes the region size. (C) Switching to a different hierarchy level, the "green regions" are collapsed, while the "blue regions" are extended. (D) Outgoing structural connectivity on a different hierarchy level where the region of the volume of interest is split into sub-regions (yellow circles).

Exploration of higher-order connectivity: To explore higher-order connectivity - the source/target regions of source/target regions - *Aggregation Queries* can be used iteratively. Therefore, volumes of interests can be selected in the query results (i.e., voxel-level connectivity of a previously executed *Aggregation Query*). This "connectivity cascade" can be performed repeatedly to explore the subsequent connectivity. We showcased this by reproducing a neuronal circuit related to learning and memory by traversing structural connectivity within the Hippocampus.

Comparison of different connectivities: Comparing the connectivities of different modalities is essential for identifying neural circuits, as they describe relations on a structural, functional, or genetic level. Previously discussed methods allow neuroscientists the retrieval and visualization of global and local connectivity on different anatomical levels. To compare them we chose a combined approach, divided into a quantitative comparison via *Connectivity Profiles* (Figure 3.12A) and a qualitative comparison as combined graph (Figure 3.12B and C). We showed in a user study that neuroscientists were intuitively able to interpret this kind of visualization.

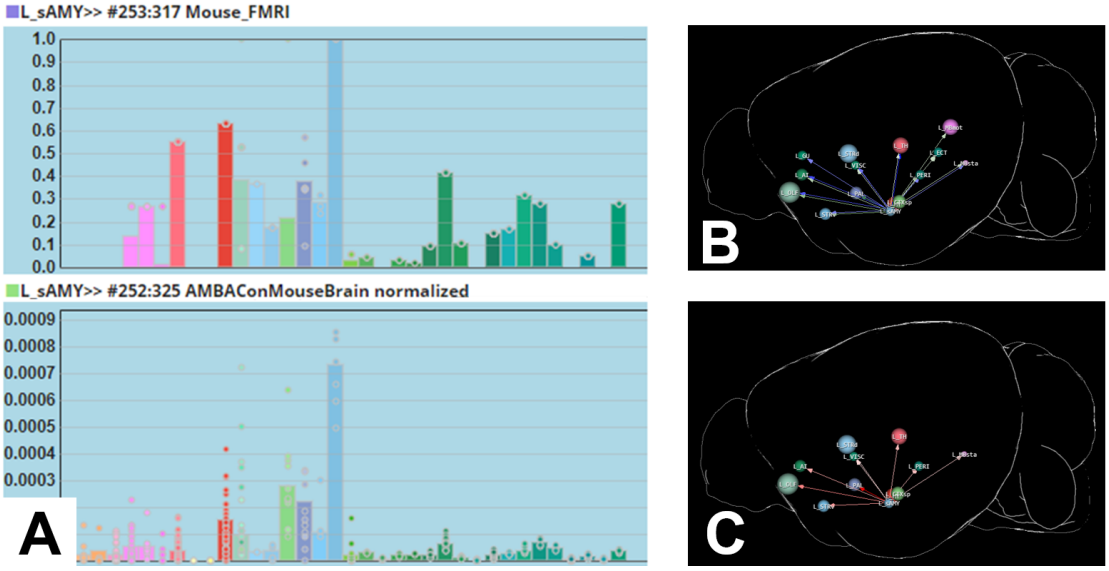


Figure 3.12: Connectivity comparison on a region-level: (A) Quantitative connectivity comparison of fMRI connectivity and structural connectivity with *Connectivity Profiles*. The bars show the mean connectivity within brain regions. Circles represent sub-regions of the respective region/bar. (B) Connectivity comparison visualizing fMRI as blue arrows and structural connectivity in green. Region colors correspond to the *Connectivity Profile*. (C) Overlap (product) of both connectivities visualized with red arrows.

Genetic dissection of networks: Exploring the genetic diversity of a brain network is relevant to determine the genetic impact on its function and consequently on behavior, meaning to know which genes are relevant in different parts of a network. Therefore, one can analyze the distribution of spatial gene expression in brain regions that are

represented by network nodes. To perform this analysis in an unbiased way (i.e., not focusing on a specific, pre-selected set of genes), one has to do this on a genome-wide scale. For this purpose, we integrated spatial gene expression data of 20,000 genes of the Allen Mouse Brain Atlas [LHA+07] into our framework presented in Paper D. We used an approach similar to *Aggregation Queries*: The user selects a volume of interest (which is part of the network) and queries the spatial gene expression data to retrieve a list of genes ordered by their mean gene expression. To achieve real-time query performance, we used spatial indexing similar to Bruckner et al. [BSG+09]. Multiple queries, and, therefore, multiple parts of a network can be compared in a parallel coordinate system that is linked to a list showing the coordinate system's selection. We demonstrated in a case study that the genetic composition of a well-known relationship in the mesolimbic system can be reproduced.

Enabling collaborative work in a web-based framework: The visual analytics methods and workflows presented in Paper D are available in a web-based, task-driven framework called *BrainTrawler*. This enables the visualization, exploration,and integration of spatial data at different scales on a consumer level PC without prior installation. The current state of the analysis can be encoded as URL (Uniform Resource Locator) in the base64 format. Saving the link allows neuroscientists the storage of their analysis, while sharing the link among colleagues may improve collaboration in a neuroscientific environment.

CHAPTER 4

Conclusion

This thesis describes a novel visual analytics approach to explore heterogeneous spatial big brain data. Chapter 1, gave an overview and a comprehensive narrative of the papers that are part of this doctoral thesis. The background of the thesis, namely the vast resources of spatial brain data and the current state-of-the-art research on how to explore this data, was discussed in Chapter 2. This step was essential for putting the individual contributions of the papers (see Chapter 3) in the appropriate scientific context.

In Paper A, we introduced a novel computational method to integrate genetic, gene expression, and connectomic information from big brain initiatives for the prediction of functional neuroanatomy. We found that functionally-related genes are not distributed at random, but are assembled into specific brain networks, which recapitulates functional neuroanatomy from literature or fMRI data. The fact that these predictions improved when incorporating higher-order network measures might reflect that the functional impact of local gene expression manifests through higher-order circuit interactions. By merging molecular genetic and structural levels of brain organization, our method has the potential to refine the functional parcellation of the brain.

In Paper B we presented a data structure for handling large brain networks on different anatomical scales in real-time, which was a prerequisite for developing an interactive visual analytics framework in the Papers C and D. By harnessing the hierarchical brain parcellation and spatial organization of the data, the technique outperformed state-of-the art graph engines in querying connectivity from network graphs with billions of edges. It further enables the retrieval of connectivity data at different resolutions, such as mesoscale structural connectivity and region-level functional connectivity on a common reference space. We demonstrated in a case study that this process can reproduce findings in neuronal circuit research. On an inter-species level, the comparison of brain networks linked to autism indicated, that there is even potential use in studying psychiatric disorders.

Inspired by the promising insights of Paper A, on how the combination of spatial genetic data and brain connectivity can predict functional neuroanatomy, we created *BrainTrawler* a novel web-based framework for analyzing and fusing heterogeneous brain data, a concept presented in Paper C and extended in Paper D. By incorporating data from large brain initiatives, *BrainTrawler* enables neuroscientists to explore genetic and functional characteristics of microcircuits without time-consuming manual data aggregation and literature research. It provides visual analytics workflows for the iterative exploration of higher-order connectivity in multimodal brain networks and on different scales, enabled by incorporating the data structure presented in Paper B. Spatial indexing of vast gene expression collections enables, for the first time, real-time dissection of brain networks genetically on a voxel-level (i.e., which genes are expressed in different parts of the networks). Furthermore, researchers can share the actions and queries they performed as persistent web-links, with visualizations familiar to neuroscientists to support collaboration and data provenance in a scientific environment. The practical relevance of this framework has been evaluated by exploring the social behavior and the memory/learning related functional neuroanatomy in mice.

4.1 Impact

This thesis might have several implications for both basic and biomedical research. We created a discovery framework that utilizes data from current popular large-scale genetic and brain network initiatives to rapidly screen for neural circuitry underlying specific brain functions, behavior, or psychiatric diseases at comparably low computing costs. The computational screening complements and may direct subsequent circuit-genetic experiments such as electrophysiology, opto-, and pharmacogenetics. If performed at a large scale with genes related to a specific behavior, our approach has the potential to refine the functional organization of the brain beyond simple anatomical domains. Importantly, our methodology could be applied to other model organisms for which spatial gene expression, networks, and genetic information is, or will be, available, for example drosophila or zebrafish.

Apart from these (rather hypothetical) implications on neuroscientific research, there were immediate effects on the involved research groups at *VRVis* and the *IMP*. First of all, the *Biomedical Image Informatics Group* at *VRVis* is engaged with data processing of different organisms - mouse, human, zebrafish, drosophila, drosophila larva, etc. - in various projects. These projects have certain synergies, so there are several technologies, components, and visualizations that can be shared among them. The methods developed during this thesis are or will be utilized for different species/projects. Furthermore, it spawned several master/bachelor theses - supported/advised by the author of this doctoral thesis - that are closely related to it or might influence future projects:

- **Dimensionality Reduction for Analysis and Visualization of Functional Connectivity in the Developing Human Brain.** Master's thesis at the Medical University of Vienna by Lisa Frauenstein in 2017 [Fra17]. The master thesis was driven by the idea of comparing different connectivities on a lower-dimensional space. This method may be included in *BrainTrawler* in future projects.
- **Guided Data Cleansing of Large Connectivity Matrices.** Master's thesis at the Technical University of Vienna by Florence Gutekunst in 2018/2019 [Gut19]. In this thesis, a tool for data cleansing of large connectivity matrices was developed. Data cleansing can be used as preprocessing for the data structure presented in Paper B. By threshold connectivity matrices and merging similar rows/columns, this tool leads to higher compression rates in the data structure, and therefore faster queries without losing too much information. This may lead to performance improvements in future projects.
- **3D Network Visualization Design Study.** Ongoing master's thesis by Daria Kruzhinskaia. The aim of this study is to develop intuitive visualizations for 3D networks beyond current practices in neuroscience. This usually involves node-link diagrams that cause problems with obfuscating edges and nodes due to many connections, which makes these visualizations hard to interpret.
- **Data-Driven Anatomical Layouting of Brain Network Graphs.** Bachelor's thesis by Gwendolyn Rippberger [Rip19]. The goal of this project is to create anatomy-driven layouts for 2D visualizations of brain networks that would enhance 2D graphs with an anatomical context to provide neuroscientists with spatial information.

At the *Haubensak Group* (*IMP*), the proposed methods are used in parts of current experimental workflows. This has led to several posters at neuroscience conferences (e.g., Federation of European Neuroscience Societies Forum 2016/2018 [KGP$^+$16, FSG$^+$18, GTF$^+$18, GPB$^+$18b] and the Goettingen Neuroscience Meeting 2017/2019 [GKP$^+$17, KGHB19]) as well as journal papers (e.g. [KGG$^+$19]) that are currently work in progress and will therefore not be discussed in this thesis.

4.2 Future Work

For the future, we aim to extend *BrainTrawler* in a holistic way. Several ideas and prototypes of projects/studies in the context of this thesis (Chapter 4.1) shall be included. Data import/preprocessing will allow experts to easily include their experimental data to directly explore them in respect to big brain initiatives in the web. Optimized 3D network visualization and 2D network visualization with anatomy-driven layouts may provide novel points of view on complex networks in their spatial context. Computing as well as visualizing network measures can provide neuroscientists with additional quantitative information.

Furthermore, the script-based GWCA for predicting functional maps (Paper A) should be fully integrated into the framework to make it more accessible to programming-illiterate neuroscientists. Moreover, an integration of large-scale multigenic behavior data - gene sets that are related to a certain behavior/function - could be used to create a comprehensive atlas of functional neuroanatomy. Publicly available and explorable as a resource in *BrainTrawler*, GWCA has the potential to extend the knowledge of functional, molecular, genetic, and structural brain organization beyond its current scale.

PAPER A

Predicting functional neuroanatomical maps from fusing brain networks with genetic information

Florian Ganglberger, Joanna Kaczanowska, Josef M. Penninger, Andreas Hess, Katja Bühler, and Wulf Haubensak. Predicting functional neuroanatomical maps from fusing brain networks with genetic information. *NeuroImage*, 2018

NeuroImage 170 (2018) 113–120

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Predicting functional neuroanatomical maps from fusing brain networks with genetic information

Florian Ganglberger [a], Joanna Kaczanowska [b], Josef M. Penninger [c], Andreas Hess [d], Katja Bühler [a,**], Wulf Haubensak [b,*]

[a] VRVis Research Center, Donau-City Strasse 11, 1220, Vienna, Austria
[b] Research Institute of Molecular Pathology (IMP), Vienna Biocenter (VBC), Campus-Vienna-Biocenter 1, 1030, Vienna, Austria
[c] Institute of Molecular Biotechnology of the Austrian Academy of Sciences (IMBA), Vienna Biocenter (VBC), 1030, Vienna, Austria
[d] Institute of Experimental and Clinical Pharmacology and Toxicology, Friedrich-Alexander University Erlangen-Nuremberg, Fahrstrasse 17, 91054, Erlangen, Germany

ARTICLE INFO

Article history:
Received 21 October 2016
Received in revised form
21 August 2017
Accepted 24 August 2017
Available online 4 September 2017

Keywords:
Neuroanatomical maps
Behavior
Connectivity
Functional neuroanatomy
Computational analysis

ABSTRACT

Functional neuroanatomical maps provide a mesoscale reference framework for studies from molecular to systems neuroscience and psychiatry. The underlying structure-function relationships are typically derived from functional manipulations or imaging approaches. Although highly informative, these are experimentally costly. The increasing amount of publicly available brain and genetic data offers a rich source that could be mined to address this problem computationally. Here, we developed an algorithm that fuses gene expression and connectivity data with functional genetic meta data and exploits cumulative effects to derive neuroanatomical maps related to multi-genic functions. We validated the approach by using public available mouse and human data. The generated neuroanatomical maps recapture known functional anatomical annotations from literature and functional MRI data. When applied to multi-genic meta data from mouse quantitative trait loci (QTL) studies and human neuropsychiatric databases, this method predicted known functional maps underlying behavioral or psychiatric traits. Taken together, genetically weighted connectivity analysis (GWCA) allows for high throughput functional exploration of brain anatomy *in silico*. It maps functional genetic associations onto brain circuitry for refining functional neuroanatomy, or identifying trait-associated brain circuitry, from genetic data.

© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

The wealth of data from brain initiatives and the increasing amount of functional genetic information creates opportunities to mine these resources for insights into the genetic and neuronal organization of brain function and behavior. Recent studies correlated brain gene expression maps with structural information to enhance our understanding of genetic and anatomical parcellation of the brain (French and Pavlidis, 2011; French et al., 2011) and its functional networks (Richiardi and Altmann, 2015; Vértes et al., 2016). These studies have been used, for instance, to explore development and physiological regulation of structural connectivity and extract functional networks *in silico* (Supplementary Note 2). Collectively, these results suggest that functional genetic information, brain gene expression data and connectomes can be successfully used for functional exploration of the brain (Supplementary Fig. 1).

Here, we mined these resources to generate functional neuroanatomical maps underlying a given brain function from genetic data. Currently, such functionally related maps are built experimentally from functional manipulations and imaging studies which are frequently invasive and costly. Discovery tools that give easily achievable and testable computational predictions would provide an ideal complementary approach into this problem.

A major challenge in this regard is that brain functions are controlled by the interaction of multiple genes within the brain. In consequence, computational predictions should reflect those functional synergies. Most established approaches that map genetic information to brain data compare gene co-expression correlation of functionally grouped genes with structural connectivity (Rubinov et al., 2015; Whitaker et al., 2016; French et al., 2011; Richiardi and Altmann, 2015). Correlative analysis of gene co-expression dissects brain organization based on the similarities of regional gene

* Corresponding author.
** Corresponding author.
E-mail addresses: buehler@vrvis.at (K. Bühler), wulf.haubensak@imp.ac.at (W. Haubensak).

https://doi.org/10.1016/j.neuroimage.2017.08.070
1053-8119/© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

expressions (Supplementary Note 2). It primarily reflects transcriptomic similarities, globally or for subsets of genes, but it is not tailored to directly infer functional synergies accumulating over multiple functionally related genes.

Motivated by this methodological gap, we sought to develop algorithms fusing sets of functionally related genes with brain data (gene expression and connectivity) to realize a semi-automatic functional parcellation from functional genetic data *in silico*. In difference to the existing correlative approaches, genetically weighted connectivity analysis (GWCA) is based on the hypothesis that functional synergies of gene sets are reflected in their cumulative impact (weights) on higher order features of fundamental brain networks such as global structural connectivity or functional resting state networks.

In the context of graph-theoretical analyses, we demonstrate that calculating the effects of cumulative gene expression on incoming/outgoing node strength generates meaningful results for functional neuroanatomy of multi-genic brain functions. When applied to gene sets from genome wide association studies (GWAS), quantitative trait loci (QTL) analyses or neurogenetic databases, these calculations allowed exploring brain circuits underlying complex behavioral traits in mice and human. Further, our workflow produced putative effector network nodes at voxel/grid level, provided as functional brain maps, that allowed a further refinement of known functional neuroanatomy.

Moreover, the GWCA framework introduced enables high throughput screening and exploration of maps of functional neuroanatomy related to gene sets. Our method is universal, digesting gene sets e.g. derived from literature meta-analyses and genetic databases, gene expression data and structural or functional connectivity data retrieved from publicly available data repositories and/or own experiments. The computed functional anatomical maps can be used for further experimental validation and refinement of neuronal circuitry underlying a specific brain function. Moreover, they allow mapping neuronal substrates affected by genetic variance linked to mental diseases with yet unknown neuronal pathophysiology (with e.g. gene associations in GWAS studies as input).

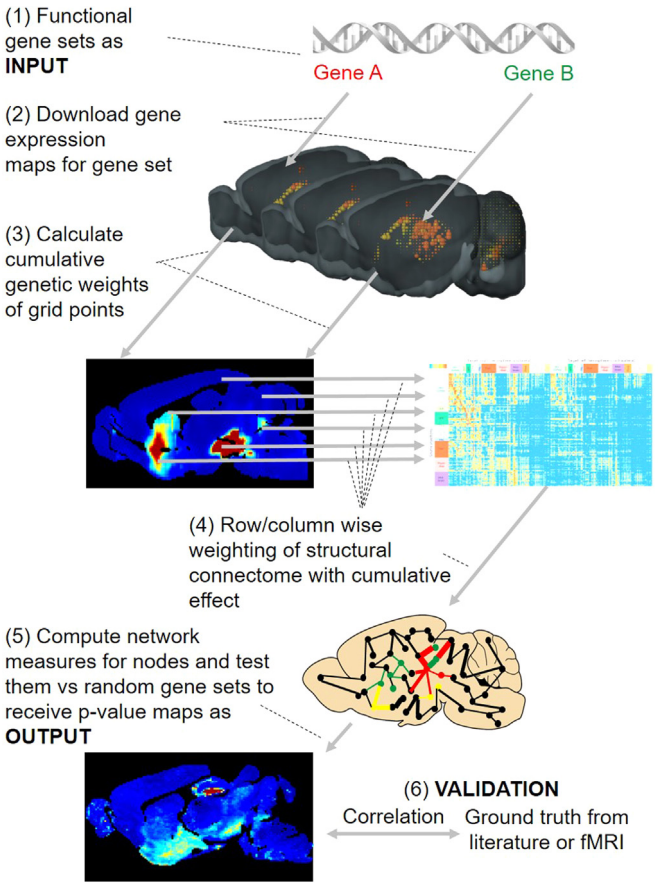


Fig. 1. Computational workflow. A functionally related gene set (e.g. social bond genes) serves as input (1). For this gene set, gene expression data is retrieved (2), normalized and used to calculate a cumulative genetic effect (3). The cumulative effect is used to weight a structural connectivity matrix (4). On the weighted network, network measures are computed and statistically evaluated by Z-tests against a null distribution (network measures based on random gene sets) (5). The output is a voxel-wise p-value map for every network measure. The results can be evaluated by computing correlation with ground truth from literature or fMRI (6).

2. Material and methods

2.1. Overview

The GWCA employs genetic-functional associations as input for weighting brain network data. This includes a set of genes associated with a given brain function or behavior, brain wide spatial gene expression data of a substantial part of the genome, and a fundamental brain network (connectome), all at the same resolution and aligned to the same common reference space. Thus, voxel, resp. grid positions and network nodes are directly corresponding.

Here, the GWCA was exemplarily based on data from the Allen Mouse Brain Atlas (AMBA), currently the most advanced template for integrated network studies of mammalian brains with extensive gene expression and connectomic information available (Lein et al., 2007; Oh et al., 2014). Of note, the method as such is universal and can be applied straight forward to data from any other species. We demonstrate this generality by applying the same method to human data using the Human Brain Atlas gene expression data framework (Hawrylycz et al., 2012). The code has been optimized for low cost parallel computing, providing a quick and convenient tool for functional exploration of brain circuitry.

Fig. 1 gives a principal overview of the workflow. After retrieving spatial gene expression data for a given input gene set (Fig. 1 (1, 2)), we are applying statistical standardization methods to compensate for spatial bias of individual gene expression and image artifacts. Based on the corrected values we calculate the *gene expression synergy* for each voxel/grid point reflecting its cumulative gene expression characteristic of the given gene set (Fig. 1 (3)). To explore its effect on the connectome, respectively source and target sites of voxel/grid points with high synergy, network data and synergy are now fused by weighting edges with related *gene expression synergy* values (Fig. 1 (4)).

To capture higher order synergies of genes spatially correlated with the underlying – now weighted - network we calculate for each node in the network, i.e. each grid point in the standard brain space, higher order network measures (Fig. 1 (5)). We compared these to network measures computed on random drawn gene sets to reveal which nodes are functionally related to the brain function or behavior associated with the input set. With this step, we inflate the gene-dependent maps to maximize the retrieval of full circuits.

2.2. Method description

We define the input set **T** of genes out of a genome-wide set **G** (Fig. 1 (1)). Gene selection strategies are described in Supplementary Note 1 Subsection 6.

If using AMBA as basis for the calculations, the spatial brain gene expression data is imported pre-aligned to a common reference space on a 100 μm grid.

The gene expression data related to **T** and **G** consists of ordered lists of gene expression densities (Lee et al., 2008) for a set of n spatial grid positions $\mathbf{p}_i \in \mathbb{R}^3 i = 1, \ldots, n$ and are stored as gene

expression density matrices **D**(**T**) and **D**(**G**) (Fig. 1 (2)).

Gene expression density is not location invariant. For example, cortical and thalamic areas in AMBA have a higher mean gene expression density than the rest of the brain. Spatial bias introduced by this variance is compensated by the standardization (Z-Score) of **D(T)** genome-wide, such that expression density distributions at every spatial position are standard-normal distributed over **G.** Results are gene and space normalized gene expression density values.

We define the synergy s_i of gene set **T** at each voxel/grid position $\mathbf{p}_i$ as the trimmed mean for the normalized gene expression density from the previous step (Fig. 1 (3)) resulting in the synergy vector $\mathbf{S} = (s_1, \ldots, s_n)$. Trimming reduced sampling artifacts in gene density maps, like image artifacts that appear as outliers with high density scores (e.g. air bubbles) (Bindhu et al., 2013).

The functional relation between genes and neuroanatomy is established by weighting either incoming or outgoing connections of every grid position by its related synergy. This highlights connections that are either targeting or are targeted by network nodes with high synergy. Given a directed connectome as the connectivity matrix $\mathbf{C} \in \mathbf{R}^{n \times n}$ (where rows represent source regions, and columns target regions on the grid positions), an incoming-synergy weighted or outgoing-synergy weighted connectome is defined as the row-respectively column-wise multiplication of **C** by **S** (Fig. 1 (4)).

To account for higher order effects on the network, i.e. where the synergy influences the properties of the network, we calculated local network measures (Rubinov and Sporns, 2010) in the weighted connectomes (Fig. 1 (5)), such as incoming/outgoing node strength, hubs, authorities, closeness, betweenness and eigencentrality on both incoming and outcoming weighted connectomes.

For statistical evaluation, we compared the grid position-wise node strength measures to randomly drawn gene sets (n = 1,000) from the genome-wide set **G** by Z-tests (Fig. 1). We adjusted the False Discovery Rate (FDR) of the p-values with the Benjamini-Hochberg (Benjamini and Hochberg, 1995) method. The results in this paper are all significant under a FDR $< 5\%$ (unless indicated otherwise).

Ultimately, these operations generated spatially distributed FDR corrected p-value maps, i.e., one p-value for every grid position.

In principle p-value maps can be computed for each combination of weighted directed connectome and network measure. Each calculated p-value map has its own interpretation. P-value maps of complementary network measures such as incoming & outgoing node strength or hubs & authorities can be fused to take into account different aspects of network context, e.g. sources and targets sites of ligand-receptor systems.

As a final step, these node-wise (for every voxel/grid position) p-value maps can be visualized in the context of the connectome to ultimately reveal the neuronal circuitry associated with the function/behavior of the input gene set (Fig. 2*A*). At first, we combined the p-value maps of the calculated network measures by using the minimum p-value for every voxel/grid position.

Our experiments demonstrated that combining incoming/outgoing node strength performed best on predicting our test data (see Result Sec. 3.1 and Fig. 2*B*). Details on the implementation for these network measures can be found in Supplementary Note 1 Sec. 3.3.

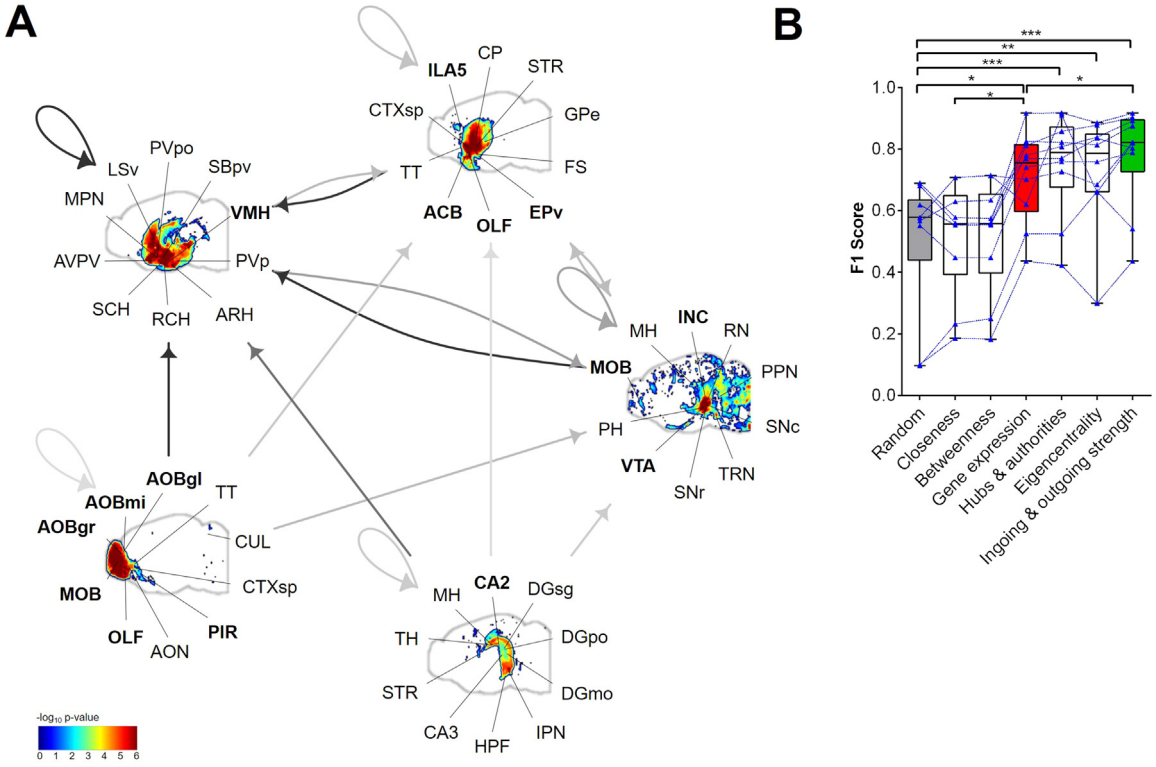


Fig. 2. Recovery of known functional anatomy from test gene sets. (*A*) Clustered nodes of a functional anatomical map (color scale indicates significance) associated with a gene set for social behavior, overlayed with structural connectivity (grey arrows, intensity indicates strength of connectivity). Loops indicate within node connections. The top-ranked networks with a known link to social behavior include main (MOB) and accessory (AOB) olfactory bulb, olfactory areas (OLF), endopiriform nucleus (EP), piriform area (PIR) and infralimbic cortex (ILA), hypothalamic nuclei (e. g., ventromedial hypothalamic nucleus (VMH)), hippocampus (particularly CA2 region), periaqueductal grey (INC – interstitial nucleus of Cajal), ventral tegmental area (VTA), and nucleus accumbens (ACB). For a complete list of abbreviations see Supplementary Tab. 1. (*B*) Node-wise comparison of predicted maps to ground truth for 10 test sets. F_1-scores increase from random classification to expression sites (gene expression synergy) and to second order network measures significantly (Benjamini & Hochberg corrected One-way ANOVA on ranks, ingoing & outgoing network strength vs expression sites; $p < 0.05$, ingoing & outgoing network strength vs random; $p < 0.001$, expression sites vs random; $p < 0.05$, eigencentrality vs random; $p < 0.01$, hubs & authorities vs random; $p < 0.05$). To compare the performance of individual sets, they are connected by blue dotted lines between network measures. Boxes indicate median and interquartile range. Incoming & outgoing node strength, hubs & authorities, closeness, betweenness and eigencentrality were tested, node strength showed the highest F_1 score.

To show all connections from the input network between every significant voxel/grid position of the combined p-value map, we reduced complexity by grouping voxels/grid positions by their connectivity, i.e. we group voxels with similar connectivity (correlation coefficient of their connections) together. The clusters are visualized by plotting a sagittally-projected heatmap of their combined p-value, surrounded by region labels. The connectivity between clusters is shown as arrows with the sum of connectivity (normalized by cluster size) given as grey-scale (Fig. 2*A*).

Additional information about the method (mathematical description, data integration, figure generation, code availability, technical resources and statistics) can be found in Supplementary Note 1.

3. Results

3.1. Proof of concept and optimization

To assess if this computational approach allows identifying function-specific brain circuitry, we focused on several well-studied gene sets, for which functional associations and functional neuroanatomy are comprehensively documented: genes associated with dopaminergic signaling, social behavior, feeding, hypothalamic–pituitary–adrenal (HPA) stress axis and synaptic plasticity. With these gene sets, we recaptured known functional neuroanatomy from literature (Supplementary Data 5).

For instance, genes associated with social behavior recapitulated their known functional neuroanatomy (Fig. 2*A*, Supplementary Data 1) (Kim et al., 2015; O'Connell and Hofmann, 2011; Young and Wang, 2004; Young et al., 2005; Leshan and Pfaff, 2014; Marlin et al., 2015; O'Connell and Hofmann, 2012).

Similarly, we were able to pick up the functional neuroanatomy (Supplementary Data 3 Case 1–10*A,B,C*, Supplementary Data 1) for other functionally-associated gene sets (Supplementary Data 3 Case 1-10*D*) including dopamine (DA) signaling, which revealed the classical DA reward VTA-ACB pathway and also motor-related connections like SN-GP (Russo and Nestler, 2013; Lammel et al., 2014; Bjoerklund and Dunnett, 2007; Berridge and Kringelbach, 2015).

The method allowed detecting the known feeding-related neuroanatomy based on genes associated with feeding, like orexin, neuropeptide Y (NPY), Agouti related protein (AgRP), proopiomelanocortin (POMC), melanocortin or leptin receptors (Betley et al., 2013; Jennings et al., 2013; Hardaway et al., 2015; Wu et al., 2012).

Different stress and fear/anxiety-related genes accumulate in the HPA axis, areas involved in control and regulation of stress and brain regions involved in processing fear/anxiety (Stoppel et al., 2006; Tovote et al., 2015; Herman and Cullinan, 1997; Smith and Vale, 2006; Carhuatanta et al., 2014; Steimer, 2002).

We also investigated gene sets for synaptic plasticity, learning and memory. As expected, these genes highlight major sites of functional and behavioral plasticity in the brain (e.g., cortex, hippocampus, amygdala) (Ressler et al., 2002; Mineur et al., 2004; Toyoda et al., 2011; Pisani et al., 2005; Lee, 2014; Hasan et al., 2013; Kirkwood and Bear, 1995; Todd and Bucci, 2015; Castro-Alamancos et al., 1995; Iriki et al., 1989).

To assess these predictions quantitatively, we collected the ground truth in form of network nodes representing regions functionally associated with these 10 gene sets from literature (Supplementary Data 2). We calculated the F_1-score (Van Rijsbergen, 1979) of precision and recall for a binary classification of the ordered voxel-wise p-values. We used this with first order network measures (expression site; genetic weight at the node itself; gene expression synergy tested voxel/grid point-wise to random drawn gene sets) and second order network measures (incoming & outgoing node strength from/to nodes with accumulated genetic weight, as well as hub score, authority score, closeness, betweenness, and eigencentrality) (Fig. 2*B*). The computational predictions correlated significantly with the known functional neuroanatomy from literature (Fig. 2*B*, green box), indicating that GWCAs assemble meaningful functional neuroanatomical maps from genetic data.

For expression sites, hubs & authorities, eigencentrality and ingoing & outgoing node strength, the F1 scores were significantly better than random (Fig. 2*B*, grey box). The predictive power increased from first order measures (Fig. 2*B*, red box) to second order measures (Fig. 2*B*, green box). This indicates that second order network measures detected regions not identified by gene expression synergy alone, yet are integrated within the same neuroanatomical map. Results for node strength showed that the prediction accuracy was superior to other network measures, and is therefore sufficient for further analysis. Importantly, the GWCA was calculated at 100 μm voxel resolution, free from *a priori* constraints from anatomical annotations and fully compatible with small rodent MRI. Thus, it is suitable to refine structure-function relationships beyond neuroanatomical scales and has the potential to identify additional nodes and subdivisions within predefined anatomical regions with possible distinct physiological functions.

To further support our findings, we overlayed computed functional maps with those obtained experimentally with fMRI. Important in the context of this paper, pain data offers the possibility to link genetics with actual fMRI (Hess et al., 2007; Hess et al., 2011; Heindl-Erdmann et al., 2010) in mice. In fact, for the pain-related gene sets (Supplementary Note 3, Supplementary Table 3 and Supplementary Data 3 Case 11-15d), the *in silico* predicted functional maps in mouse brain were reproducing large portions of the functional neuroanatomy observed with Blood-Oxygen-Level-Dependent functional magnetic resonance imaging (BOLD fMRI data, warped onto the AMBA reference space by optimized ANTS (Avants et al., 2008) parametrization) *in vivo* (Fig. 3*A* and *B*). This further substantiates the validity of our approach. While our method seemed to fit best with sets of >4 genes (Supplementary Fig. 2), predictions were also informative at the single-gene level. Functional imaging data of Cacna2d3 mutants, a highly conserved pain gene, revealed altered thalamo-cortical connectivity and synesthesia after thermal stimulation in mutant mice (Neely et al., 2010). The predicted maps computed from Cacna2d3 alone (Fig. 3*A*, top right) recaptured pain functional neuroanatomy from fMRI (Fig. 3*A*, bottom left, 3*B*) and pain maps that are affected by this gene (Fig. 3*A*, bottom right, Fig. 3*B*). Nevertheless, the single gene operations will depend heavily on the gene itself, and so we recommend using gene sets for the most efficient and accurate functional neuroanatomy integration.

Taken together, these data show that GWCAs of functionally related gene sets generate meaningful functional neuroanatomy of multi-genic brain functions. If applied to many iterations of functionally grouped genes, this type of approach could be extended for the semi-automated functional annotation and parcellation of the brain *in silico*.

3.2. Functional maps from multi-genic meta data

A central aim in basic neuroscience and psychiatry is to understand how genetic variations control behavioral traits. One of the challenges is that behavioral traits, as brain functions, are largely multi-genic. Therefore, identifying the neural circuitry through which these trait-associated genes contribute to phenotypic differences is experimentally hard. Based on these results, we applied GWCA to explore yet unknown or only partially described brain circuitry underlying behavioral traits investigated in genetic screens or association studies. We expanded our analysis on pain and included fear/anxiety and autism spectrum disorder (ASD)

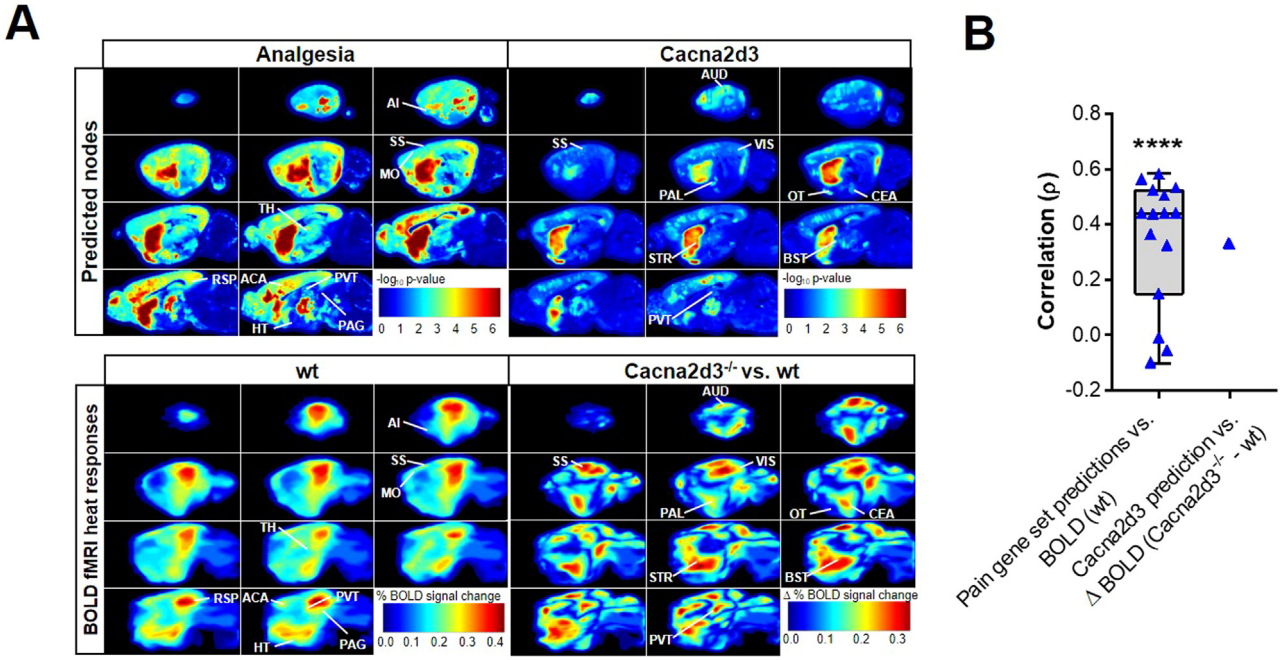


Fig. 3. Computed functional maps correlate with BOLD fMRI. (*A*) Similarity of functional maps nodes predicted for analgesia gene sets and Cacna2d3 gene (*top*) to nodes with heat evoked fMRI responses (*bottom*). (*Left*) Of the highest ranked predicted nodes for the Analgesia gene set, insular (AI), anterior cingulate (ACA), somatosensory (SS), motor (MO) and retrosplenial (RSP) cortices, periaqueductal grey (PAG), thalamic (TH, PVT) and hypothalamic (HY) areas correspond to prominent nodes in fMRI. (*Right*) Of the highest ranked predicted nodes for the Cacna2d3 gene, striatum (STR), paraventricular nucleus of thalamus (PVT), bed nuclei of stria terminalis (BST), pallidum (PAL), central amygdalar nucleus (CEA), sensory cortices (somatosensory areas (SS), visual areas (VIS), auditory areas (AUD)) and olfactory tubercle (OT) and correspond to those identified by fMRI. Color bars indicate $-\log_{10}$-scaled p-values (*top*) and heat stimulus responses (% BOLD signal changes) in wt animals (*bottom left*) or differences (Δ) in heat responses between Cacna2d3$^{-/-}$ and wt animals (% BOLD signal changes in Cacna2d3$^{-/-}$ - % BOLD signal changes in wt animals) (*bottom right*). For a detailed list of brain regions see Supplementary Table 1. (*B*) Voxel-wise Spearman correlations of p-value maps predicted from pain gene sets with BOLD fMRI responses. The box indicates median and interquartile range of Spearman's ρ. Blue triangles indicate the correlation of the individual pain gene sets (top to bottom: analgesia, synaptic signaling, hypersensitivity, memory, response to drug, cognition, learning or memory, calcium ion transport, calcium signaling, nociception, neurological system process, regulation of neurological system process, response to pain and response to stress) to wild-type BOLD signal.

gene sets (Supplementary Note 3) from publicly available databases and published meta-studies (Supplementary Table 3). In some cases, large gene sets were clustered using the DAVID (Huang et al., 2009) platform to parcellate them into functional category-linked subsets, and so in those cases genes are not only related by the analyzed trait, but also regarding sub-functions annotated in the database. When supplied with these gene sets, the GWCA extracted meaningful functional maps (Supplementary Data 3 Case 11–29). These maps, of which node-wise comparisons are in line with their functional annotation from literature, give a comprehensive representation of functional genetic synergies underlying the respective trait (Fig. 4*A*, green squares).

Extending our approach to human template based on resting state networks from fMRI (as reference brain network) demonstrated that the methodology can be generalized to other species (Fig. 4*B*). Cross-validation with the meta-studies (Supplementary Data 4, Supplementary Table 2) reveals similar findings for both species (Fig. 4*A* and *B*), demonstrating its versatility for functional exploration of the human brain in health and disease *in silico*.

Of note, within the mouse and human framework, the algorithm also identified nodes not yet linked to the query trait, thereby extracting potential novel elements (Fig. 4*A* and *B*, blue squares) of functional brain networks (see Discussion for details).

4. Discussion

4.1. Integrating genetic, gene expression and connectomic information

We have shown that GWCA successfully integrates genetic, gene expression and connectomic information from brain and genomic initiatives for rapid functional exploration of the brain *in silico*. We found that, in the brain, functionally related genes are not distributed at random but assemble into specific maps, which recapitulate functional anatomical annotations and/or functional data from fMRI. Cumulative effects, from expression sites alone (Fig. 2*B*, red bar), reflect functional synergies within functionally related genes, which are not directly fitted by transcriptomic similarities, usually derived from correlative analysis (Supplementary Note 2). GWCA predictions further improved by second order network measures, which incorporate functional synergies of local gene expression that manifest in the context of higher-order interactions within the brain architecture.

Incoming & outgoing node strength (Fig. 2*B*, green bar) performed best, but not significantly better than hubs & authorities or eigencentrality. In contrast to these network measures covering the influence of nodes on networks, betweenness and closeness highlight the effect of shortest paths in networks (Rubinov and Sporns, 2010). They outlined small distinctive nodes, that are part of functional properties, but failed to predict the entirety of functional neuroanatomy (explaining the seemingly random F_1-score in Fig. 2*B*). According to Watts and Strogatz (1998), the influence of path length measures systematically decreases with increasing randomness of a network, whereas cluster measures remain high over a much wider range of randomness. Therefore, this finding supports the notion that GWCA captures real non-random small-world networks found in the brain (Watts and Strogatz, 1998).

Moreover, the superior performance by incoming & outgoing node strength may be explained by the fact this measure captures best the cumulative genetic effects emerging directly within network nodes and their primary connections through direct genetic

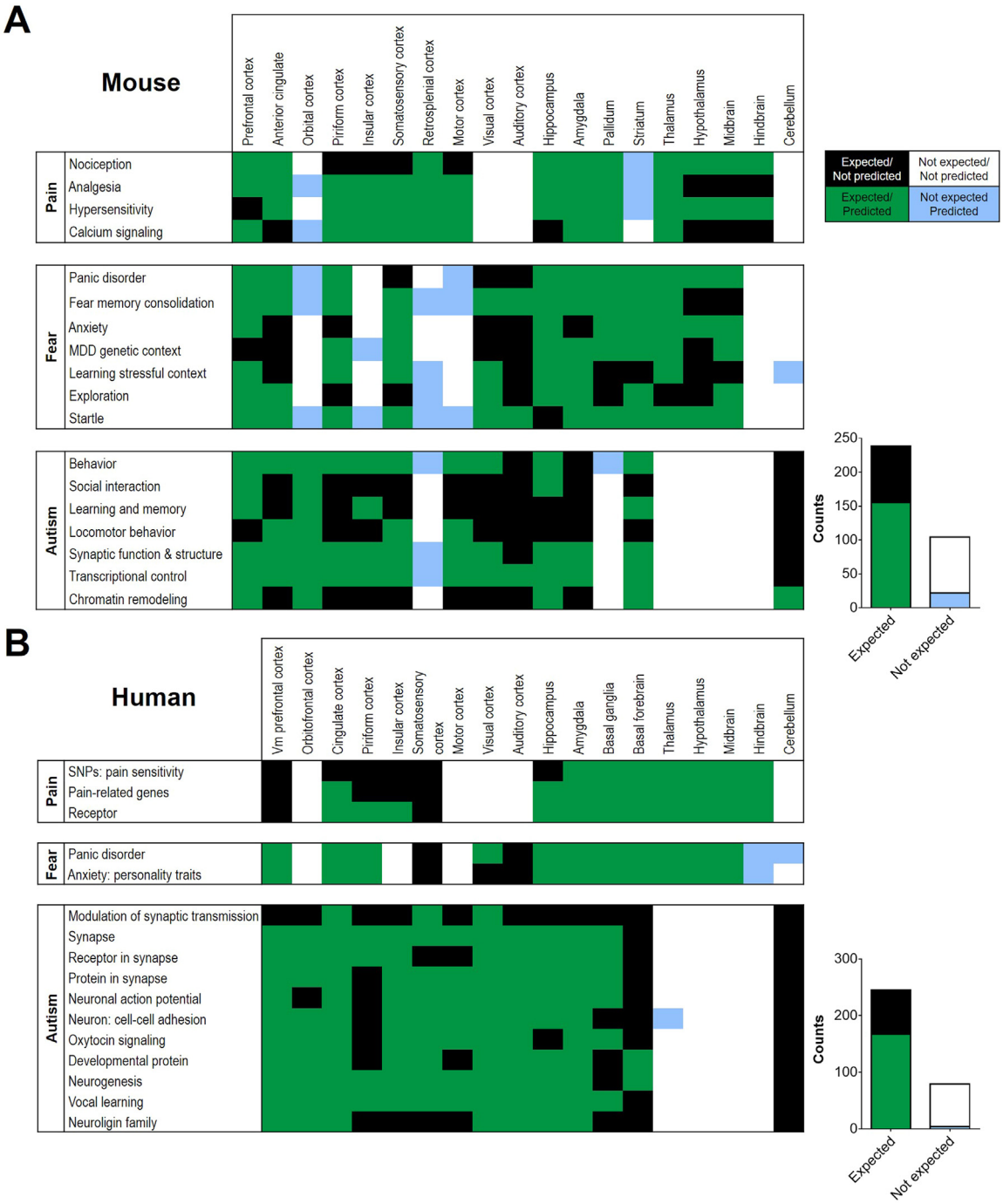


Fig. 4. Predicting functional maps of behavioral traits from mouse and human genetic meta data. (*A*) Left, node-wise comparison of predicted mouse functional anatomy for pain, fear and autism, divided into different functional subcategories, to functional neuroanatomical annotations from literature for the top p-value ranked nodes. Right, Quantification of the qualitative assessment. There is a significant overlap between predicted maps and functional neuroanatomical annotation ($n = 342$; Fisher's exact test, $p < 0.0001$). (*B*) Left, node-wise comparison of predicted human functional anatomy for pain, fear and autism, divided into different functional subcategories, to functional neuroanatomical annotations from literature for the top 100 p-value ranked nodes. Right, quantification of the qualitative assessment. There is a significant overlap between predicted maps and functional neuroanatomical annotation ($n = 324$; Fisher's exact test, $p < 0.0001$).

(molecular) interactions. In addition, this measure naturally reflects that gene expression sampled here takes place in local somata, which builds up the nodes and sources/target sites thereof. In contrast, path-centric measures like betweenness and closeness might be better suited for inspecting effects on network routing which may reflect complex features of higher order dynamic states in the context of specific neuronal activity, but not the anatomical node driven network structure influenced by gene expression.

Taken together, by fusing cumulative gene expression and best-fit network measures, we provide an optimized tool that derives meaningful functional neuroanatomical maps from genetic information.

4.2. *Refining functional anatomical annotations*

When applied to gene sets from behavioral genetics, we demonstrated that our workflow can extract putative effector network nodes as functional brain maps which can be used to explore

trait-specific circuitries. These explorations allowed us to refine several known functional neuroanatomy (Fig. 4, green squares). For instance, the anatomy of thalamo-cortical and cortico-cortical connections in thermal pain processing can be dissected to fine anatomical resolution (e.g., Supplementary Data 3 Case 11*E*, red arrows, note layer specificity) which could not be achieved with fMRI (Fig. 3*A*, wt). GWCA, when based on startle response QTLs, extracted a specific and strong connection between PVT and central amygdala (Supplementary Data 3 Case 22*E*, red arrow). Interestingly this connection recently emerged as central element in fear control (Do-Monte et al., 2015; Penzo et al., 2015). Similarly, for ASD, we identified many cortico-cortical connections (Supplementary Data 3 Case 23-29*E*, red arrows) with prediction accuracy reaching individual layers. Among similar lines, the method uncovered circuitry within regions functionally not yet commonly associated with the respective trait: for instance, the functional association of visual cortex with pain processing (Supplementary Data 3 Case 13, 14, 15*E*, Supplementary Data 1) (Gopalakrishnan et al., 2015), motor cortex with startle response (Supplementary Data 3 Case *22E*) (Kühn et al., 2004) and hypothalamic circuitry with autism (Supplementary Data 3 Case 27*E*, Supplementary Data 1) (Kurth et al., 2011). This can be particularly useful to link genetic variance and neurophysiology in mental diseases with unknown etiopathology (with e.g., gene associations from GWAS studies as input).

5. Conclusion

GWCA significantly adds to the understanding of structure-functional relationships for several reasons. First, it allows for generating functional neuroanatomical maps from genetic data. Second, when performed iteratively with multiple functionally grouped gene sets at larger scales, this allows to genetically define functional parcellation of the brain. Third, when applied to functional gene sets from meta-studies or behavioral trait analysis, it allows to rank order brain circuits according to their role in that given function or behavioral trait. These candidate circuits can then serve as entry points for further functional validation, e.g., with opto- and pharmacogenetic methods.

The functional relation underlying our study can be exploited to associate gene sets with specific brain functions or brain functions with specified gene sets (Supplementary Fig. 1). Importantly, our strategy applies to other neural systems (beyond mouse and human) for which genetic information, gene expression maps and connectomes are, or will be, available and allows exploration of functional brain organization in cases where actual functional data is difficult, if not impossible, to obtain.

Taken together, GWCA emerges as a timely tool for mining genetic and brain initiatives for insight into the genetic and functional organization of the brain and mind. This study highlights synergies that emerge from fusing data across different platforms and should spark discussions about similar strategies in the future.

Author contributions

F.G., J.K. and W.H. conceived the method. F.G implemented the method, performed data analysis and the quantitative validation. J. K performed the qualitative validation. J.P. and A.H. provided fMRI data. F.G., J.K., A.H., K. B. and W.H. wrote the manuscript. K.B. and W.H. jointly supervised the project.

Competing financial interests

The authors declare no competing financial interests.

Acknowledgments

W. H. was supported by a grant from the European Community's Seventh Framwork Programme (FP/2007–2013)/ ERC grant agreement no. 311701, the Research Institute of Molecular Pathology (IMP), Boehringer Ingelheim and the Austrian Research Promotion Agency (FFG) (Headquarter 852936). A. H. was supported by the BMBF NeuroRad (02NUK034D) and BMBF NeuroImpa (01EC1403C).

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.neuroimage.2017.08.070.

References

Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C., 2008. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Med. Image Anal. 12 (1), 26–41. http://dx.doi.org/10.1016/j.media.2007.06.004.

Benjamini, Yoav, Hochberg, Yosef, 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B Methodol. . http://dx.doi.org/10.2307/2346101

Berridge, Kent C., Kringelbach, Morten L., 2015. Pleasure Systems in the Brain. Neuron. Cell Press.

Betley, J. Nicholas, Cao, Zhen Fang Huang, Ritola, Kimberly D., Sternson, Scott M., 2013. Parallel, redundant circuit organization for homeostatic control of feeding behavior. Cell 155 (6), 1337–1350.

Bindhu, Pr, Krishnapillai, Rekha, Thomas, Priya, Jayanthi, P., 2013. Facts in artifacts. J. Oral Maxillofac. Pathol. JOMFP 17 (3), 397–401 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3927342&tool=pmcentrez&rendertype=abstract.

Bjoerklund, Anders, Dunnett, Stephen B., 2007. Dopamine neuron systems in the brain: an update. Trends Neurosci.

Carhuatanta, Kimberly A.K., Shea, Chloe J.A., Herman, James P., Jankord, Ryan, 2014. Unique genetic loci identified for emotional behavior in control and chronic stress conditions. Front. Behav. Neurosci. 8, 341 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4204525&tool=pmcentrez&rendertype=abstract.

Castro-Alamancos, M.A., Donoghue, J.P., Connors, B.W., 1995. Different forms of synaptic plasticity in somatosensory and motor areas of the neocortex. J. Neurosci. Offic. J. Soc. Neurosci. 15 (July), 5324–5333.

Do-Monte, Fabricio H., Quinones-Laracuente, Kelvin, Quirk, Gregory J., 2015. A temporal shift in the circuits mediating retrieval of fear memory. Nature 519 (7544), 460–463. http://dx.doi.org/10.1038/nature14030, Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.

French, L., Tan, P.P., Pavlidis, P., 2011. Large-scale analysis of gene expression and connectivity in the rodent brain: insights through data integration. Front. Neuroinf. 5, 12. http://dx.doi.org/10.3389/fninf.2011.00012.

French, Leon, Pavlidis, Paul, 2011. Relationships between gene expression and brain wiring in the adult rodent brain. PLoS Comput. Biol. 7 (1).

Gopalakrishnan, R., Burgess, R.C., Plow, E.B., Floden, D.P., Machado, A.G., 2015. A magnetoencephalography study of multi-modal processing of pain anticipation in primary sensory cortices. Neuroscience 304 (September), 176–189. http://dx.doi.org/10.1016/j.neuroscience.2015.07.049.

Hardaway, J.A., Crowley, N.A., Bulik, C.M., Kash, T.L., 2015. Integrated circuits and molecular components for stress and feeding: implications for eating disorders. Genes Brain Behav. 14 (1), 85–97. http://dx.doi.org/10.1111/gbb.12185.

Hasan, Mazahir T., Samuel, Hernández-González, Dogbevia, Godwin, Treviño, Mario, Bertocchi, Ilaria, Gruart, Agnès, Delgado-García, José M., 2013. Role of motor cortex NMDA receptors in learning-dependent synaptic plasticity of behaving mice. Nat. Comm. 4, 2258 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3759079&tool=pmcentrez&rendertype=abstract.

Hawrylycz, Michael J., Lein, Ed S., Guillozet-Bongaarts, Angela L., Shen, Elaine H., Ng, Lydia, Miller, Jeremy A., van de Lagemaat, Louie N., et al., 2012. An anatomically comprehensive Atlas of the adult human brain transcriptome. Nature 489 (7416), 391–399. http://dx.doi.org/10.1038/nature11405, Nature Publishing Group.

Heindl-Erdmann, Cornelia, Axmann, Roland, Kreitz, Silke, Zwerina, Jochen, Penninger, Andreas, Schett, Georg, Brune, Kay, Hess, Andreas, 2010. Combining

functional magnetic resonance imaging with mouse genomics: new options in pain Research. Neuroreport 21 (1), 29–33 http://www.ncbi.nlm.nih.gov/pubmed/19934782.
Herman, James P., Cullinan, William E., 1997. Neurocircuitry of stress: central control of the hypothalamo-pituitary-adrenocortical Axis. Trends Neurosci.
Hess, A., Axmann, R., Rech, J., Finzel, S., Heindl, C., Kreitz, S., Sergeeva, M., et al., 2011. Blockade of TNF-alpha rapidly inhibits pain responses in the central nervous system. Proc. Natl. Acad. Sci. U. S. A. 108 (9), 3731–3736. http://dx.doi.org/10.1073/pnas.1011774108.
Hess, Andreas, Sergejeva, Marina, Budinsky, Lubos, Ulrich Zeilhofer, Hanns, Brune, Kay, 2007. Imaging of hyperalgesia in rats by functional MRI. Eur. J. Pain 11 (1), 109–119.
Huang, Da Wei, Lempicki, Richard A., Sherman, Brad T., 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4 (1), 44–57.
Iriki, A., Pavlides, C., Keller, A., Asanuma, H., 1989. Long-term potentiation in the motor cortex. Science 245 (4924), 1385–1387. http://dx.doi.org/10.1126/science.2551038.
Jennings, Joshua H., Rizzi, Giorgio, Stamatakis, Alice M., Ung, Randall L., Stuber, Garret D., 2013. The inhibitory circuit architecture of the lateral hypothalamus orchestrates feeding. Science (New York, N.Y.) 341 (6153), 1517–1521 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4131546&tool=pmcentrez&rendertype=abstract.
Kim, Yongsoo, Venkataraju, Kannan Umadevi, Pradhan, Kith, Mende, Carolin, Taranda, Julian, Turaga, Srinivas C., Arganda-Carreras, Ignacio, et al., 2015. Mapping social behavior-induced brain activation at cellular resolution in the mouse. Cell Rep. 10 (2), 292–305, Elsevier.
Kirkwood, A., Bear, M.F., 1995. Elementary forms of synaptic plasticity in the visual cortex. Biol. Res. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=8728822
Kühn, Andrea A., Sharott, Andrew, Trottenberg, Thomas, Kupsch, Andreas, Brown, Peter, 2004. Motor cortex inhibition induced by acoustic stimulation. Exp. Brain Res. 158 (1), 120–124. http://dx.doi.org/10.1007/s00221-004-1883-4.
Kurth, Florian, Narr, Katherine L., Woods, Roger P., O'Neill, Joseph, Alger, Jeffry R., Caplan, Rochelle, McCracken, James T., Toga, Arthur W., Levitt, Jennifer G., 2011. Diminished gray matter within the hypothalamus in autism disorder: a potential link to hormonal effects? Biol. Psychiatr. 70 (3), 278–282. http://dx.doi.org/10.1016/j.biopsych.2011.03.026.
Lammel, S., Tye, K.M., Warden, M.R., 2014. Progress in understanding mood disorders: optogenetic dissection of neural circuits. Genes Brain Behav. 13 (1), 38–51.
Lee, Chang-Kyu, Sunkin, Susan M., Kuan, Chihchau, Thompson, Carol L., Pathak, Sayan, Ng, Lydia, Lau, Chris, et al., 2008. Quantitative methods for genome-scale analysis of in situ hybridization and correlation with microarray data. Genome Biol. 9 (1), R23 http://genomebiology.com/2008/9/1/R23.
Lee, Yong-Seok, 2014. Genes and signaling pathways involved in memory enhancement in mutant mice. Mol. Brain 7 (1), 43 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4050447&tool=pmcentrez&rendertype=abstract.
Lein, Ed S., Hawrylycz, Michael J., Ao, Nancy, Ayres, Mikael, Bensinger, Amy, Bernard, Amy, Boe, Andrew F., et al., 2007. Genome-wide Atlas of gene expression in the adult mouse brain. Nature 445 (7124), 168–176. http://dx.doi.org/10.1038/nature05453.
Leshan, Rebecca L., Pfaff, Donald W., 2014. The hypothalamic ventral premammillary nucleus: a key site in Leptin's regulation of reproduction. J. Chem. Neuroanat. (Elsevier)
Marlin, Bianca J., Mitre, Mariela, D'amour, James A., Chao, Moses V., Froemke, Robert C., 2015. Oxytocin enables maternal behavior by balancing cortical inhibition. Nature 520 (7548), 499–504 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4409554&tool=pmcentrez&rendertype=abstract.
Mineur, Y.S., Crusio, W.E., Sluyter, F., 2004. Genetic dissection of learning and memory in mice. Neural Plast. 11 (3–4), 217–240 http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15656270.
Neely, G. Gregory, Hess, Andreas, Costigan, Michael, Keene, Alex C., Goulas, Spyros, Langeslag, Michiel, Griffin, Robert S., et al., 2010. A genome-wide Drosophila screen for heat nociception identifies α2δ3 as an evolutionarily conserved pain gene. Cell 143 (4), 628–638.
O'Connell, L.A., Hofmann, H.A., 2012. Evolution of a vertebrate social decision-making network. Science 336 (6085), 1154–1157.
O'Connell, Lauren A., Hofmann, Hans A., 2011. The vertebrate mesolimbic reward system and social behavior network: a comparative synthesis. J. Comp. Neurol.
Oh, Seung Wook, Harris, Julie A., Ng, Lydia, Winslow, Brent, Cain, Nicholas, Mihalas, Stefan, Wang, Quanxin, et al., 2014. A mesoscale connectome of the mouse brain. Nature 508 (7495), 207–214 http://www.nature.com/doifinder/10.1038/nature13186.
Penzo, Mario A., Robert, Vincent, Tucciarone, Jason, Dimitri, De Bundel, Wang, Minghui, Van Aelst, Linda, Darvas, Martin, et al., 2015. The paraventricular thalamus controls a central amygdala fear circuit. Nature 519 (7544), 455–459. http://dx.doi.org/10.1038/nature13978, Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.
Pisani, Antonio, Centonze, Diego, Bernardi, Giorgio, Calabresi, Paolo, 2005. Striatal synaptic plasticity: implications for motor learning and Parkinson's disease. Mov. Disord.
Ressler, Kerry J., Paschall, Gayla, Zhou, Xiao-liu, Davis, Michael, 2002. Regulation of synaptic plasticity genes during consolidation of fear conditioning. J. Neurosci. Off. J. Soc. Neurosci. 22 (18), 7892–7902.
Richiardi, Jonas, Altmann, Andre, 2015. Correlated gene expression supports synchronous activity in brain networks. Science 348 (6240), 11–14.
Rubinov, Mikail, Sporns, Olaf, 2010. Complex network measures of brain connectivity: uses and interpretations. NeuroImage 52 (3), 1059–1069.
Rubinov, Mikail, Ypma, Rolf J.F., Watson, Charles, Bullmore, Edward T., 2015. Wiring cost and topological participation of the mouse brain connectome. Proc. Natl. Acad. Sci. 112 (32), 201420315 http://www.pnas.org/content/early/2015/07/24/1420315112.abstract.html?etoc.
Russo, Scott J., Nestler, Eric J., 2013. The brain reward circuitry in mood disorders. Nat. Rev. Neurosci. 14 (9), 609–625 http://www.nature.com/nrn/journal/v14/n9/full/nrn3381.html%5Cn http://www.nature.com/nrn/journal/v14/n9/pdf/nrn3381.pdf.
Smith, Sean M., Vale, Wylie W., 2006. The role of the hypothalamic-pituitary-adrenal Axis in neuroendocrine responses to stress. Dialogues Clin. Neurosci.
Steimer, Thierry, 2002. The biology of fear- and anxiety-related behaviors. Dialogues Clin. Neurosci.
Stoppel, C., Albrecht, A., Pape, H.C., Stork, O., 2006. Genes and neurons: molecular insights to fear and anxiety. Genes Brain Behav.
Todd, Travis P., Bucci, David J., 2015. Retrosplenial cortex and long-term memory: molecules to behavior. Neural Plast. (Hindawi Publishing Corporation)
Tovote, Philip, Paul Fadok, Jonathan, Lüthi, Andreas, 2015. Neuronal circuits for fear and anxiety. Nat. Rev. Neurosci. 16 (6), 317–331 http://www.ncbi.nlm.nih.gov/pubmed/25991441.
Toyoda, Hiroki, Yao Li, Xiang, Jun Wu, Long, Gao Zhao, Ming, Descalzi, Giannina, Chen, Tao, Koga, Kohei, Zhuo, Min, 2011. Interplay of amygdala and cingulate plasticity in emotional fear. Neural Plast. (Hindawi Publishing Corporation)
Van Rijsbergen, C.J., 1979. Information Retrieval, second ed. Butterworths.
Vértes, Petra E., Rittman, Timothy, Whitaker, Kirstie J., Romero-Garcia, Rafael, Váša, František, Kitzbichler, Manfred G., Wagstyl, Konrad, et al., 2016. Gene transcription profiles associated with inter-modular hubs and connection distance in human functional magnetic resonance imaging networks. Philosophical Trans. R. Soc. B Biol. Sci. . http://dx.doi.org/10.1098/rstb.2015.0362
Watts, Duncan J., Strogatz, Steven H., 1998. Collectivedynamics of 'Small-World' networks. Nature 393 (6684), 440–442. http://dx.doi.org/10.1038/30918.
Whitaker, Kirstie J., Vértes, Petra E., Romero-Garcia, Rafael, Váša, František, Moutoussis, Michael, Prabhu, Gita, Weiskopf, Nikolaus, et al., 2016. Adolescence is associated with genomically patterned consolidation of the hubs of the human brain connectome. Proc. Natl. Acad. Sci. 113 (32), 9105–9110. http://dx.doi.org/10.1073/pnas.1601745113.
Wu, Qi, Clark, Michael S., Palmiter, Richard D., 2012. Deciphering a neuronal circuit that mediates appetite. Nature 483 (7391), 594–597.
Young, L.J., Wang, Z., 2004. The neurobiology of pair bonding. Nat. Neurosci. 7 (10), 1048–1054 http://www.ncbi.nlm.nih.gov/pubmed/15452576.
Young, Larry J., Young, A.Z.M., Hammock, E.A.D., 2005. Anatomy and neurochemistry of the pair bond. J. Comp. Neurol. 493, 51–57.

Supplementary Information

Supplementary Figures 1-2.

Supplementary Data 1. P-values of first and second order effects for all cases based on region (mouse and human).

Supplementary Data 2. Ground truth generated from literature.

Supplementary Data 3. Functional neuroanatomical maps, significant regions and network visualization of all cases used in this paper for mouse.

Supplementary Data 4. **S**ignificant regions of all cases used in this paper for human.

Supplementary Data 5. Gene sets used in this paper

Supplementary Table 1. Anatomical abbreviations.

Supplementary Note 1

1. Mouse Data

The mouse connectome was retrieved as (structural) connectivity from all 2173 available injection sites (state March 2016) to their target sites given as image data, detailing projections labeled by rAAV tracers via serial two-photon tomography (Oh et al. 2014). Those sites are added up to a connectivity matrix which covers about 15 percent of the right hemisphere as source regions, and about 100% as target regions. The AMBA connectome (right hemisphere injections) was mirrored onto (left hemisphere) AMBA gene expression data. In order to also take weak connections into account, the connectome was binarized by a threshold according to Oh, S. W. *et al.* (Oh et al. 2014), Extended Data Figure 7, that minimizes the amount of false positive connections. The gene expression density is interpolated to a 100 micron resolution to match the resolution of the connectome. A Matlab script for downloading the gene expression for **T** and for **G**, as well as the AMBA connectome is provided on request.

2. Human data

Gene expression by region retrieved from the Allen Human Brain Atlas (Hawrylycz et al. 2012). The Allen Institute provides an affine transformation to MNI152 (Fonov et al. 2011) space by its API. We used resting state functional connectivity from the Human Connectome Project (Glasser et al. 2013), which is also in MNI152space (Fonov et al. 2011). Data normalization was performed in a robust way (median/mad) since fewer data points are more sensitive to outliers compared to mouse data.

3. Mathematical description

3.1 Input data:

We expect that the spatial brain gene expression data and the voxel/grid resolution connectivity data are pre-aligned to a common reference space.

We retrieve for each n grid position $\mathbf{p}_i \in \mathbb{R}^3$ i=1,…,n and each available gene g_j in the mouse genome $\mathbf{G} = \{g_j\}_{j=1..m}$ (or at least a random drawn subset) the gene expression density vector

$$d_i(\mathbf{G}) = (d_{i1},\ldots,d_{im}) \quad | \quad i=1,\ldots,n$$

and store it as gene expression density matrix

$$\mathbf{D(G)} = (d_1(\mathbf{G}),\ldots,d_n(\mathbf{G}))^T{}_{i=1..n} = (d_{ij})_{\,i=1..n\ j=1..m} \in \mathbb{R}^{nxm}$$

We define the input as a function/trait associated set $\mathbf{T} = \{t_k\}_{k=1,\ldots,l}$ of genes out of a genome-wide set **G**, resulting in the expression density matrix $\mathbf{D(T)} \in \mathbb{R}^{nxl}$. Respectively, $d_i(\mathbf{T})$ denotes a gene expression density vector for subset **T**.

3.2 Normalization*:*

Bias introduced by spatial variance in gene expression density is compensated by the standardization (Z-Score) of **D(T)** genome-wide, such that expression density distributions at every spatial position are standard-normal distributed over **G:**

$$d_{ik}^{\text{gene normalized}} = (d_{ik} - \mu_i)/\ \sigma_i \quad | \quad \forall\ d_{ik} \in \mathbf{D(T)}$$

where $\mu_i = \mu(d_i(\mathbf{G}))$ and $\sigma_i = \sigma(d_i(\mathbf{G}))$.

Subsequently, standardization is performed for $\mathbf{D}^{\textbf{gene normalized}}(T) = (\ d_{ij}^{\textbf{gene normalized}})$ in their spatial distribution pattern to adjust for differences between genes within the overall brain expression density:

$$d_{ik}^{\text{gene-space normalized}} = (d_{ik}^{\text{gene normalized}} - \mu_j{}^{\text{gene normalized}})/\ \sigma_j \quad | \quad \forall\ d_{ik}^{\text{gene normalized}} \in \mathbf{D}^{\text{gene normalized}}(\mathbf{T})$$

where $\mu_j^{\text{gene normalized}} = \mu\ (d_i^{\text{gene normalized}}(\mathbf{T}))$ and $\sigma_j^{\text{gene normalized}} = \sigma\ (d_i^{\text{gene normalized}}(\mathbf{T}))$.

For the AMBA data set we replaced missing values with 0 (which is the most likely value that a value can have after normalization in genome space) for the calculation of μ_j and σ_j to compensate for missing lateral slices.

We define the gene expression synergies $s_i(\mathbf{T})$, i =1,…,n at a grid positions $\mathbf{p}_i$, i=1,…,n for gene set **T** as the trimmed means of the gene-space normalized gene expression density values for all genes in set **T.**

$$s_i = \mu_{\text{trimmed}}(d_i^{\text{gene-space normalized}}(\mathbf{T})) \quad | \quad i=1,\ldots,n$$

The functional relation between genes and neuroanatomy is expressed by weighting either incoming or outgoing connections of every grid position by **S**. Given the directed AMBA connectome as a connectivity matrix

$$\mathbf{C} = (c_{vw})_{v,w = 1..n} \; ; \mathbf{C} \in R^{\,n \times n}$$

where the rows represent the source regions, the columns target regions, either an incoming $\mathbf{C}^{\mathbf{weighted\ in}}$ or outgoing $\mathbf{C}^{\mathbf{weighted\ out}}$ weighted directed connectome is defined as

$$\mathbf{C}^{\mathbf{weighted\ out}} = (c_{vw}^{\text{weighted out}})_{w= 1,\ldots,n} = s_v * (c_{vw})_{w = 1,\ldots,n} \quad | \quad \forall\, v = 1..n$$

$$\mathbf{C}^{\mathbf{weighted\ in}} = (c_{vw}^{\text{weighted in}})_{v = 1,\ldots,n} = s_w * (c_{vw})_{v = 1,\ldots,n} \quad | \quad \forall\, w = 1..n$$

3.3 Effect calculation:

To account for higher order synergies within functional maps, we computed those maps from local network measures (Rubinov and Sporns 2010) of the weighted connectomes $\mathbf{C}^{\text{weighted in}}$ and $\mathbf{C}^{\text{weighted out}}$. The incoming node strength (sum of incoming connections for every node) of $\mathbf{C}^{\text{weighted in}}$ and $\mathbf{C}^{\text{weighted out}}$ is defined as

$$\mathbf{IN}^{\text{weighted out}} = in_v^{\text{weighted out}} = \sum_{w=1}^{n} c_{vw}^{\text{weighted out}} \mid \forall\, v = 1..n$$

$$\mathbf{IN}^{\text{weighted in}} = in_v^{\text{weighted in}} = \sum_{w=1}^{n} c_{vw}^{\text{weighted in}} \mid \forall\, v = 1..n$$

and the outgoing node strength (sum of outgoing connections for every node) as

$$\mathbf{OUT}^{\text{weighted out}} = out_w^{\text{weighted out}} = \sum_{v=1}^{n} c_{vw}^{\text{weighted out}} \mid \forall\, w = 1..n$$

$$\mathbf{OUT}^{\text{weighted in}} = out_w^{\text{weighted in}} = \sum_{v=1}^{n} c_{vw}^{\text{weighted in}} \mid \forall\, w = 1..n$$

3.4 Statistical evaluation:

We compared the position-wise node strength measures to randomly drawn gene sets (n=1000) from the genome-wide set **G** by Z-tests and adjusted the False Discovery Rate (FDR) the Benjamini-Hochberg (Benjamini and Hochberg 1995) method.

The significance of $\mathbf{IN}^{\text{weighted out}}$ can be interpreted as nodes that are receiving from primary expression sites (regions with high **S)**, while $\mathbf{OUT}^{\text{weighted in}}$ shows regions projecting to primary expression sites. P-value calculations of $\mathbf{IN}^{\text{weighted in}}$ and $\mathbf{OUT}^{\text{weighted out}}$ are numerically equal to the p-value calculation of **S (**for a node degree>0), since for those cases the sum of incoming and outgoing connections are constant factors when compared to random effects. We point this out to clarify the p-value calculation of $\mathbf{IN}^{\text{weighted in}}$ and $\mathbf{OUT}^{\text{weighted out}}$ can be substituted by **S** for computational reasons.

$$in_v^{\text{weighted in}} = \sum_{w=1}^{n} (s_v * c_{vw}) = s_v * \sum_{w=1}^{n} c_{vw} \quad | \quad \forall\, v = 1..n$$

$$out_w^{\text{weighted out}} = \sum_{v=1}^{n} (s_w * c_{vw}) = s_w * \sum_{v=1}^{n} c_{vw} \quad | \quad \forall\, w = 1..n$$

3.5 Output:

A p-value map (a p-value for every grid position) for every effect. In this paper, **S, IN**$^{\text{weighted}}_{\text{out}}$**, OUT**$^{\text{weighted}}_{\text{in}}$ are used due to their fast computation, simplicity and biological significance.

4. Code availability

The code for retrieving data (gene expression, mouse connectome) from the AMBA API consists of a Matlab script whose single input parameter is a .csv with function/trait information as a list of gene symbols and Entrez IDs. The main algorithm was implemented as an R-script that uses the generated files (downloaded data from AMBA) of the Matlab script to normalize, calculate and carry out a statistical evaluation to generate p-value maps and structural network visualization for every testcase. The statistical evaluation, which was randomized because of the extent of the computational task, is parallelized.

MATLAB- and R-codes are publically available under an open source license (BSD License) for non-commercial use at https://github.com/NeuroscienceTools/GWCA.

5. Figure generation

Figures were generated with a R-script that will be provided on request. It uses the p-value maps of the method to generate slice-views of different effects, heatmaps with statistical measures of the effects and gene expression, clustered networks, csv-files with raw data and precision-recall heatmaps (for data with ground truth).

1. Slice-views: Slice-views show 11 maximum intensity projections of 5 sagittal slices each of a 132x80x114 voxel volume (which represents grid positions) that shows the left hemisphere of the mouse brain, interpolated (gaussian) for higher resolution images. Slice-views are used to visualize a log-scaled mapping of first order p-values (of **S**), second order incoming node strength **IN** (regions that are targets of first order regions) and second order **OUT** (regions projecting to first order regions). At the bottom-right corner is a color-bar, indicating the minus $\log_{10}$-scaled p-values, the threshold for false positive FDR (10% solid line, 5% dotted line). Slice-views of all testcases can be found in Supplementary Data 3 Case 1-30*A, B, C*.

2. Heatmaps: Heatmaps in Supplementary Data 3 Case 1-30*D* and Supplementary Data 4 show the log-scaled p-values of first and second order effects as well as single gene effects (gene expression density of a gene vs gene expression density of the genome) for every significant region (a region that has at least one voxel with significant first or second order effect). The regions are color-coded (on the left side) corresponding to the AMBA, and given by their acronym on the right side. Similar information can be found in the attached csv files (Supplementary Data 1) which contain the region-wise p-values of first and second order effects.

3. Clustered network graphs: We clustered our test sets via hierarchical clustering with Ward's Criterion (Murtagh and Legendre 2011) using the R function `hclust(*, "ward.D2")`. To ensure that voxels with similar connections are within the same cluster, they are clustered by their Pearson-correlation coefficient of their connectivity. To visualize the clusters, we plotted a sagittally-projected heatmap of their combined p-value (minimum p-

value of effects), surrounded by region labels. The connectivity between clusters is shown by the sum of connectivity (normalized by cluster size) between the clustered regions given as grey-scale. All graphs can be found in Fig. 2*A* and Supplementary Data 3 Case 1-30*E*.

4. F_1-score bar-chart (Figure 2B): Based on available ground truth from the literature (Supplementary Data 2), we calculated the F_1-score (Van Rijsbergen 1979) based on the precision and recall for a binary classification of ordered p-values. It doesn't take the true negative rate into account, which is acceptable for the following reason: The literature-based ground truth is region based. This means we can identify

- true positives (a positive classified voxel within a region of the ground truth)
- false positive (a positive classified voxel outside a region of the ground truth)

but not

- true negative (a negative classified voxel outside a region of the ground truth), since the total set of regions of the functional neuroanatomy are still unknown
- false negatives (a negative classified voxel within the ground truth), since it is possible that only a subset of the ground truth region is specific for functional neuroanatomy.

For the calculation of the F_1-score, respectively precision and recall, the precision is computed as the ratio of true positive voxels to the amount of positive voxels. For a voxel-based recall, a false negative rate would be necessary, and so we used the region-based recall, the ratio of positive classified regions to ground truth regions. We defined a positive classfied region if at least 5% of the voxels of a region is positive (to account for noise). P-value maps for the F_1-score bar chart were computed at 200 micron resolution due to extensive computational network measures.

5. Validation table (Figure 4): A region-wise ground truth for pain, fear and anxiety related genes sets was derived from literature (Supplementary Note 3 and Supplementary Table 3) which devides brain-regions in "expected" (true positives) and "not expected" (true negatives) regions for every gene-set. We used a binary classification of ordered p-values, analogues to the F_1-score bar-chart but with one difference: Since "not expected" represents "true negative" regions, we maximized the youden's index (Sensitivity + Specificity -1) (Youden 1950) of the prediction. The threshold for binary classification was choosen to maximize the youden's index, and limited to significant p-values (FDR<=10%), i.e. not-significant regions were always classified as negatives.

6. Gene set selection

Genes for testsets were either selected by manually compiling them by literature research on genes for which functional associations and functional neuroanatomy are comprehensively documented (Section 3.1 Figure 2B), or from multigeneic meta data (Section 3.2, Figure 4). Validation of manually compiled gene sets in Figure 2B showed a mean F1 score of 0.77 across 10 testes when compared to the ground truth. Correctly predicted voxels that are related to the functional association of their gene set were significantly (FDR<=10%) different from random drawn gene sets. In contrast, gene sets from multi-genic meta data were defined by genetic screens or association studies, and therefore did not underly a manual selection of genes according to their functional neuroanatomy. Those sets may

contain genes that do not show gene expression synergy with the others or not/ubiquitously brain expressed genes. This leads to unspecific synergies across the brain, respectively a low contrast between high and low synergy areas. Furthermore, low-contrast synergy reduces significance of their difference to random drawn gene sets.

To improve the contrast of a gene set **T** we sought to remove those genes not contributing to the synergy (see above), so that the resulting subset t_{opt} maximizes the standard deviation of the synergy across the brain.

$$t_{opt} = \underset{t \in \mathrm{P(T)}}{arg\,max}(\sigma(\mathbf{S}(t)) * \sqrt{n_t})$$

where P(**T**) is the power set of **T** (set of all subsets of **T**), $\sigma(\mathbf{S}(t))$ the sample standard deviation of the synergy across the brain (synergy of all voxel) of the gene set $t \in \mathrm{P}(\mathbf{T})$. The median absolut deviation (MAD) was used as estimation for σ to be more robust against local strong peaks in synergy. Since the synergy is defined as the trimmed mean, $\sigma(\mathbf{S}(t))$ needs to be corrected by the factor $\sqrt{n_t}$ according to the central limit theorem to adjust for different sample sizes n_t. Otherwise the sample standard deviation would depend on the sample size. To avoid maxima of small t_{opt} that do not represent the overall structure of $\mathbf{S}(t)$, we limited the optimization by

$$\rho_{\mathrm{S(t)},\mathrm{S}(t_{opt})} > \varepsilon$$

which means that the pearson correlation coefficient ρ, between the synergy $\mathbf{S}(t)$ and the synergy of the optimized set $\mathbf{S}(t_{opt})$, must be greater than a certain threshold ε. As a consequence, $\mathbf{S}(t_{opt})$will have a higher contrast than $\mathbf{S}(t)$, while showing synergy on similar regions. We found that $\varepsilon = 0.75$ (for human) and $\varepsilon = 0.95$ (for mouse) and showed the best results after validation with the youden's index on test cases used in Figure 4 (see Supplementary Note 1 Section 5.5 for details on validation).

Finding t_{opt} has been performed by using a genetic algorithm from the R package genealg (Willighagen 2015) adapted for parallel computing. To see if the genetic algorithm produces consistent results with gene sets already associated with functional neuroanatomy, we first tested it with the manual compiled literature sets. Supplementary Table 4 shows the effect of the optimization on the amount of genes in the genesets (#), the validation with ground truth (F1) and their similarity of the resulting p-value maps with the original gene-set by their spearman rank-correlation coefficient (cor). Comparing the results of the original set (org.) with the sets after optimization (after opt.) showed that in most cases, only 0 or 1 genes were sorted out, in others $\frac{1}{3}$ (especially at the larger sets Feeding and HPA Axis Central Control). Comparison with ground truth showed that the F1 score was on average equal before and after optimization. The high correlation of the p-value maps indicates that the removed genes do not have a large influence on the gene expression synergy. Further we added random genes to the optimized sets (25%, 50% and 75% of their size) to assess the stability of the optimization. It showed that the amount of genes that were left were independent of the amount of random genes (comparison of avg #opt), but the amount of random genes could be reduced (from 25% to 8,7%, 50% to 17% and 75% to 23%).

We also validated the optimization on random testsets (n=400), with different sizes (5,10,20,30). Under an FDR=10%, the mean amount of significant voxels across different p-value maps (gene expression synergy, incoming node strength, outgoing node strength) was 10% of the brain, independent of the set size. Therefore, the false positive results of random sets, optimized with the genetic algorithm, are still equal the expected error rate.

We applied gene set selection for pain, fear and several autism related gene sets for human (Supplementard Data 4 Case 1-7, 8-9, 11 and 16) since they did not show significant difference to random drawn gene sets due to low contrast. This improved the mean youden's index of Figure 4 (human) from 0,314 to 0,62. Applying the gene set selection to all test cases in Figure 4, the mean youden's index did not improve significantly (0,44 to 0,45 for mouse and 0,62 to 0,64). The original gene sets for mouse and human, as well as the optimized gene sets for Supplementary Data 4 Case 1-7, 8-9, 11 and 16 are listed in Supplementary Data 5.

Gene-set	**org.**		**after opt.**			**+25% random**	**+50% random**	**+75% random**
	#	**F1**	**#**	**F1**	**cor**	**avg #opt**	**avg #opt**	**avg #opt**
Social bonds	8	0.91	8	0.91	1	8 (14%)	7.9 (2%)	8 (27%)
Central Amygdala Microcircuitry	4	0.44	4	0.44	1	4 (14%)	4 (27%)	3.75 (32%)
Dopaminergic system	10	0.77	10	0.77	1	9.3 (6%)	9.5 (18%)	9.25 (21%)
Feeding	30	0.80	20	0.80	0.89	18.1 (8%)	17.1 (18%)	16.75 (25%)
HPA Axis Central Control	29	0.82	20	0.85	0.95	14.6 (5%)	15.1 (12%)	15 (15%)
HPA Axis regulation	16	0.82	12	0.85	0.95	10.3 (6%)	10.3 (6%)	10.5 (13%)
Hypothalamic Input to Central Amyg.	4	0.55	4	0.55	1	4 (14%)	4 (27%)	4 (29%)
Social behavior	12	0.89	12	0.89	1	11.3 (6%)	11.6 (14%)	11,5 (23%)
Long Term Potentiation	11	0.89	7	0.87	0.94	6 (1%)	6.5 (4%)	6.5 (16%)
Synaptic plasticity	10	0.83	9	0.79	0.95	8.2 (15%)	7.7 (25%)	8.25 (32%)
					Average:	**8.9 (8.7%)**	**8.9 (17%)**	**8.9 (23%)**

Supplementary Table 4. Performance of the contrast optimization on manual compiled genesets. The table shows the amount of genes (#), the F1 score calculated with ground truth (F1) and the spearman rank correlation of the p-value maps to the original gene set (cor) as well as the mean amount of original genes after optimization (avg #opt). The brackets show the percentage of random sets that are still left. The table contains the original gene set (org.), the gene set after optimization (after opt.) and the gene set after optimization with random genes added (+% random).

7. *Technical resources*

We used the Amazon elastic cloud computing service with an "r3.8xlarge" instance (32 cores, 244 GB RAM) (Amazon 2015). More than 100 GB RAM is recommended, 40 GB alone to hold the connectivity matrix in the memory. Additional memory is needed for parallel processing (approximately 3 GB per core). We tested the R-scripts with 30 cores. The

computation uses about 200 GB Ram and takes between 1 and 2 hours per testcase (depending on the amount of genes in a set) to calculate the p-values for first and second order effects. The clustering for the circle-graphs are also parallelized. Depending on the size of the significant areas, clustering takes between 30 minutes to 3 hours.

8. General statistics

Unless indicated otherwise, data were tested for normality by Kolmogorov–Smirnov or D'Agostino & Pearson tests at $\alpha<0.05$ and analyzed non-parametrically if tests didn't pass. Predicted functional neuroanatomy maps were compared to ground truth from fMRI using a Spearman correlation of the $-\log_{10}$-scaled voxel-wise p-value of predicted nodes, set to $p=10^{-3}$ for all $p<10^{-3}$, to BOLD heat responses of wt animals or differences in BOLD heat responses in Cacna2d3 mutant vs. wt animals, respectively. To compensate for registration errors between the AMBA reference space and fMRI data, these comparisons were performed on volumes downsampled to 400 μm spatial resolution.

Supplementary Note 2

Investigating functional and structural brain network data and its analysis is an ongoing challenge (Bullmore and Sporns 2009). Bullmore and Sporns (Bullmore and Sporns 2009) described the exploration of structural and functional brain networks as a multi-stage approach, beginning with the separate creation of structural and functional connectivity matrices based on anatomical parcellations. Network measures, such as Node degree, Node strength, Hubs, Authorities, Centrality, Betweenness etc., indicate network properties of interest when compared to equivalent measures of a population of random networks (null-distribution). A local (region-wise) or global (Mantel-test) (Mantel 1967) comparison reveals functional and structural correspondences of the networks.

The integration of genetic information facilitates insight into the influence on neuronal activity and structural organization of the brain (Leon French and Pavlidis 2011). French and Pavlidis (Leon French and Pavlidis 2011) compared cortical and subcortical regions of a rat connectome (Bota, Dong, and Swanson 2005) and AMBA gene expression data (Lein et al. 2007) using Spearman's rank correlation to show that brain regions with similar expression patterns have more similar connectivity profiles. The similarities are close enough that a computational model by Ji et al (Ji, Fakhry, and Deng 2014) could predict structural connectivity by gene expression profiles. 4048 genes with coronal spatial expression data were used as individual features in a sparse model to obtain a predictive accuracy of 93% on anatomical parcellations. A follow up study proved that this also works on mesoscale-resolution (voxels at 200 micron resolution) (Fakhry and Ji 2015).

A combined approach of comparing structural connectivity, gene co-expression correlation and functional networks was investigated by (Richiardi and Altmann 2015). Resting-state fMRI networks (default-mode, salience, sensorimotor and visuospatial) were used as a starting point to identify functionally related cortical regions in mice and humans. The strength fraction (scaled node strength of gene co-expression networks) between those regions was significantly more similar than to the remaining brain regions (tested by permutation tests). Genes that are related to the four functional networks were identified by

ranking them by their marginal influence on the strength fraction. A gene co-expression matrix including only top-ranked genes was compared to structural connectivity using the Mantel procedure (Mantel 1967) and were significant compared to a sample of 10,000 random gene sets. (L French, Tan, and Pavlidis 2011) used Spearman's rank correlation between node degree of structural connectivity and gene co-expression of gene sets related to Gene Ontology groups (cellular composition and biological process) to assess how structural connectivity is genetically driven. Connectivity related Gene Ontology groups were also used by Fulcher and Fornito (Fulcher and Fornito 2016). They showed that the mean gene co-expression correlation of groups related to biological processes are higher for connections involving structural "hubs" (node degree over threshold) vs non-hubs indicates topological specializations of interregional connections. Structural network hubs were also found to correspond to known functional networks from the literature (Rubinov et al. 2015; Whitaker et al. 2016). Compared to other studies (Richiardi and Altmann 2015; Fulcher and Fornito 2016; Leon French and Pavlidis 2011; L French, Tan, and Pavlidis 2011) which used node strength or variations of it, Rubinov and Sporns (Rubinov and Sporns 2010) assessed other structural network parameters, such as community structures, hierarchical modules, high-low cost sub-networks etc.

An overview of related work and its modalities can be found in Supplementary Table 1. Apart from Fakhry and Ji (Fakhry and Ji 2015), who used high-resolution prediction, the studies cited were computed on anatomically parcellated mouse brains (Richiardi and Altmann (Richiardi and Altmann 2015) also used human data). Our approach was performed on 100-micron grid parcellation. In contrast to Richiardi and Altmann (Richiardi and Altmann 2015), where functionally related gene sets were products of their marginal influence on resting-state networks, we used functionally-linked gene sets as the entry point of our method. Fulcher and Fornito, as well as French et al. (Fulcher and Fornito 2016; L French, Tan, and Pavlidis 2011) showed the influence of Gene Ontology groups of biological processes on structural networks, while our approach utilized sets from gene association studies (database-mining, QTL analyses or SNPs) and that can be directly linked to certain behavioral or mental features. Known functional networks from the literature confirmed our results as well as the correlation with resting state fMRI.

Comparing gene co-expression correlation to structural connectivity is a common approach for assessing brain structures with genetic functionality (Richiardi and Altmann 2015; Fulcher and Fornito 2016; Ji, Fakhry, and Deng 2014; Fakhry and Ji 2015; Leon French and Pavlidis 2011; L French, Tan, and Pavlidis 2011; Rubinov et al. 2015; Whitaker et al. 2016). The novelty in our paradigm is weighting structural connectivity with functionally related, cumulative gene. It is not only comparing networks, but it shows the direct effect of functionally related gene expression on brain anatomy. Those effects were encountered by node strength, which we proved to be a sufficient indicator, but also with various other network measures**.**

Supplementary Note 3

Pain sensation is biomedically one of the most important brain functions. While physiological sensation is essential to protect the organism and to avoid harm, it is very often a result of diseases or pathological/abnormal processes when the sensory information does not reflect the factual danger from the environment. Pain gene sets from mice and human sutdies were

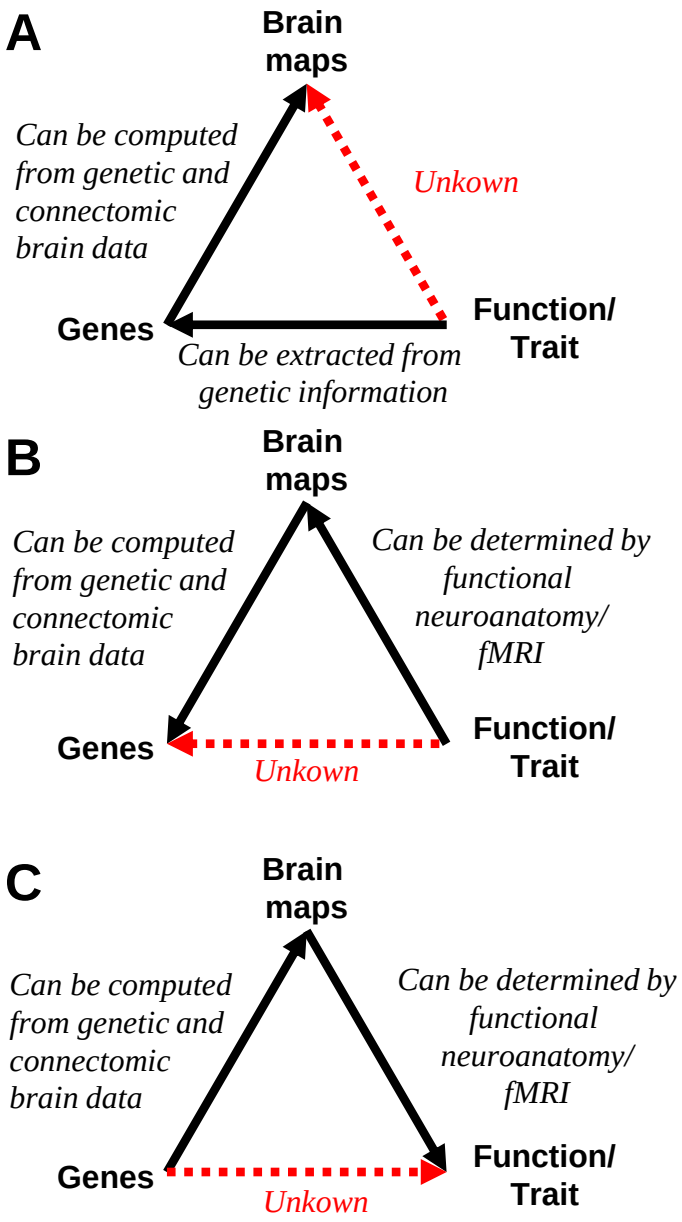
taken from literature and databases (Supplementary Table 3) (LaCroix-Fralish, Ledoux, and Mogil 2007; Neely et al. 2012, https://doi.org/10.1016/j.neuroscience.2016.04.041), pre-clustered or pre-assigned to subcategories based on behavioral phenotype (nociception, analgesia, hypersensitivity) or functional annotations (Gene Ontology: synaptic signaling, memory, response to drug, cognition, learning or memory, calcium ion transport, calcium signaling (calmodulin binding associated genes related to pain processing), neurological system process, regulation of neurological system process, response to pain and response to stress). For the human case we chose a metastudy combining SNPs associated with pain sensitivity or we extracted subcategories (obtained using the DAVID platform based on functional annotation) from the database for pain-related genes. We also used the Calcium signaling category as a set based on evolutionary conserved pain genes. Importantly, the effector networks from most of these gene sets could be linked to known pain-related areas in the brain (Tracey 2008; Denis et al. 2015; Hess et al. 2007; Heindl-Erdmann et al. 2010), but also other regions such as piriform and entorhinal cortices, nucleus accumbens and VTA (Fig. 4*A*, Supplementary Data 1). Functional neuroanatomy maps from these gene sets, and the single gene Cacna2d3, were also compared to fMRI pain responses of wt and mutant animals, respectively (Neely et al. 2010) (Fig. 3*A*). The maps derived from the gene sets were similar to the expected pain network from the mouse fMRI (Fig. 3*A*). The Cacna2d3-dependent maps identified by our method retraced Cacna2d3's functional genetic effects on pain processing in fMRI in regions like striatum, olfactory areas, somatosensory cortex, hippocampus, hypothalamus, paraventricular nucleus of thalamus (PVT) and basal ganglia. Similarly, for the human gene sets (Fig. 4*B*), we obtained the brain regions known to be involved in pain processing, including central grey, PVT, insular and somatosensory cortex, but also VTA – as in the mouse case – or higher order associative cortices which are responsible for self-awareness and conscious perception of pain.

Fear and anxiety-related genes were retrieved from JAX QTLs database (mouse) or from literature (mouse and human) (Santos, D'Amico, and Dierssen 2015; Eppig et al. 2015), pre-assigned to behavioral phenotypes (startle response, exploration, anxiety, depression and panic disorder). Again, the computed maps (mouse and human) contained nodes with a fitting functional annotation, like fear-related regions in the amygdalar complex, prefrontal cortex, thalamic or midbrain structures (Katche et al. 2013; Ferreira et al. 2003; Bradfield and McNally 2010; Garcia et al. 1999; Schoenbaum, Chiba, and Gallagher 1998; Morrison, Dias, and Ressler 2015). Moreover, the main nodes detected by our method are in line with their associated functional subcategory, e.g. startle behavior was linked to insular cortex and PVT, while mental disorders were linked to insular cortex, ACB and VTA (Fig. 4*A*). For the panic disorder category, we can see differences in cortical regions identified for mouse and human. For example, human data, unlike the mouse, lacks vmPFC, somatosensory or motor corices, while we did not detect the auditory cortex in the mouse brain (Fig. 4).

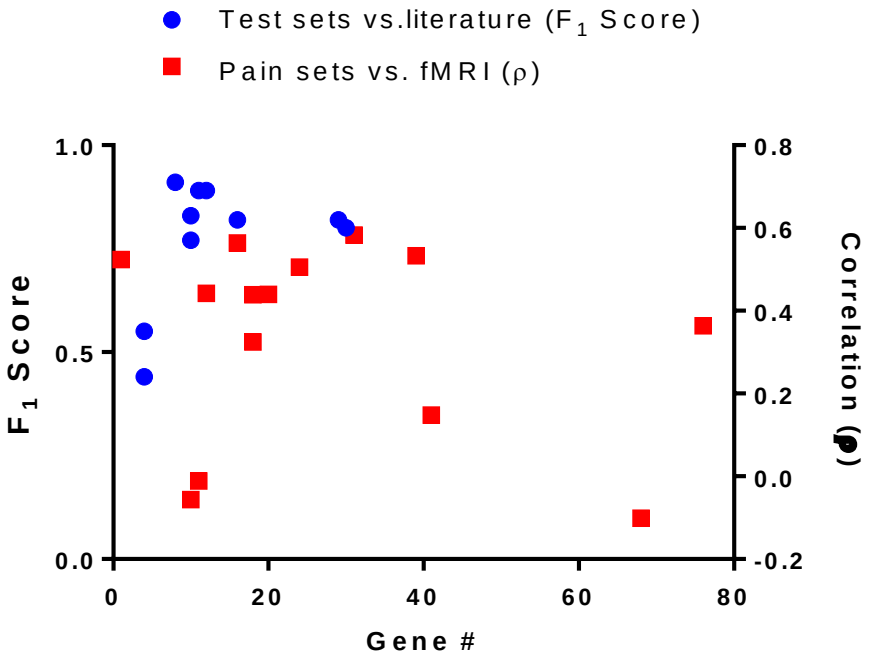
For autism-related genes, we retrieved 183 genes implicated in behavioral phenotypes in mouse models of ASD and 739 autism-associated genes in humans from Autdb database (Basu, Kollu, and Banerjee-Basu 2009) and clustered the genes with DAVID (Huang, Lempicki, and Sherman 2009), for further analysis, we chose functional annotation categories that were the most relevant for ASD modeling: linked to behavior, cognitive abilities, synaptic functions and cellular level processes. Similar to the other gene sets, the computationally predicted maps contained nodes related to autistic brain function (Kennedy and Adolphs 2012; Critchley et al. 2000; Sahin and Sur 2015; Bickart et al. 2014; McAlonan et al. 2005; Zhan et al. 2014; Bourgeron 2015; Anomal et al. 2015; Santos, D'Amico, and

Dierssen 2015), in the case of the human brain several cortical, subcortical and cerebellar areas were not identified (Fig. 4*B)*.

To sum up, we were able to identify most of the known functionally involved brain regions for all of the investigated categories based on mouse and human data. Additionally, for different specific subcategories the method identified functionally relevant structures which were found at the highest positions in rank-order lists. Taking together all the data, the method can also be a useful tool for identifying novel functional targets, potentially involved in traits linked to the genetic input. With this, we can bridge already known functional systems using potential new -still unexplored - connections or even identify new functional networks. For more detailed information please see Supplementary Data 1, 2, Fig. 3, Supplementary Data 3 Case 11-29 (for mouse) and Supplementary Data 4 (for human).



Supplementary Figure 1. Principle predictions from genetic and connectomic brain data. (*A*) Predicting functional brain maps from fusing genetic information (e. g., from genetic association studies) with genetic and connectomic brain data (this study). (*B*) Identifying genes underlying a specific behavioral trait from fusing functional brain maps (e. g., from literature, fMRI) with genetic and connectomic brain data. (*C*) Assigning a behavioral trait to a set of genes from fusing genetic and connectomic brain data with functional brain maps (e. g., from literature, fMRI).



Supplementary Figure 2. Size of gene sets and prediction reliability. The reliability of the predictions (comparison of the functional maps predicted for 10 test sets vs. ground truth from literature (Fig. 2*B*, right bar) and of the functional maps predicted for pain sets vs. actual fMRI (Fig. 3*B*)) plotted against the gene set size in these comparisons. There is no significant correlation (Spearman correlations ρ(F1 scores)=0.2577, p=0.4674, n=10 and ρ(ρ)=-0.06093, p=0.8319, n=15).

PAPER B

A data structure for real-time aggregation queries of big brain networks

Florian Ganglberger, Joanna Kaczanowska, Wulf Haubensak, and Katja Bühler. A data structure for real-time aggregation queries of big brain networks. *Neuroinformatics*, pages 1–19, 2019

“This is a post-peer-review, pre-copyedit version of an article published in NeuroInformatics. The final authenticated version is available online at: http://dx.doi.org/10.1007/s12021-019-09428-9

A Data Structure for real-time Aggregation Queries of Big Brain Networks

Florian Ganglberger[1], Joanna Kaczanowska[2], Wulf Haubensak[2], Katja Bühler[1]

[1]VRVis Research Center, Donau-City Strasse 11, 1220 Vienna, Austria

[2]Research Institute of Molecular Pathology (IMP), Vienna Biocenter (VBC), Campus-Vienna-Biocenter 1, 1030 Vienna, Austria

Abstract Recent advances in neuro-imaging allowed big brain-initiatives and consortia to create vast resources of brain data that can be mined by researchers for their individual projects. Exploring the relationship between genes, brain circuitry, and behavior is one of the key elements of neuroscience research. This requires fusion of spatial connectivity data at varying scales, such as whole brain correlated gene expression, structural and functional connectivity. With ever-increasing resolution, these tend to exceed the past state-of-the art in size and complexity by several orders of magnitude. Since current analytical workflows in neuroscience involve time-consuming manual data-aggregation, incorporating efficient techniques for handling big connectivity data is a necessity.

We propose a novel data structure enabling the interactive exploration of heterogeneous neurobiological connectivity data with billions of edges. Based on this data structure we realized *Aggregation Queries*, i.e. the aggregated connectivity from, to or between brain areas allows experts to compare the multimodal networks residing at different scales, or levels of hierarchically organized anatomical atlases. Executed on-demand on volumetric gene expression and connectivity data, they allow an interactive dissection of networks in real-time and based on their spatial context. The data structure is optimized in order to be accessible directly from the hard disk, since connectivity of large-scale networks typically exceeds the memory size of current consumer level PCs. This allows experts to embed and explore their own experimental data in the framework of public data resources without the need for their own large-scale infrastructure.

Our data structure outperforms state-of-the-art graph engines in retrieving connectivity of arbitrary user defined local brain areas. We demonstrate the feasibility of our approach by analyzing fear-related functional neuroanatomy in mice. Further, we show its versatility by comparing multimodal brain networks linked to autism. Importantly, we achieve cross-species congruence in retrieving human psychiatric traits networks, which facilitates the selection of neural substrates to be further studied in mouse models.

Introduction

Recent brain initiatives, such as the Allen Institute (Oh et al. 2014; Hawrylycz et al. 2012; Lein et al. 2007), the Human Brain Project (Markram et al. 2011), the WU-Minn Human Connectome Project (Van Essen et al. 2013), and the China Brain Project (Poo et al. 2016), have accumulated large sets of brain data for neuroscience research. Visual analytics emerges as a promising tool to mine this multimodal neurobiological data for insight into the functional organization of the brain (K. Li et al. 2012). Such technologies allow the direct exploration of relations between genes, neuronal circuitry and brain function and can quickly add context to experimental findings. However, the major challenges for visual analytic workflows arise from accessing, fusing and visualizing spatial brain data, such as brain gene expression, structural and functional connectivity, and non-spatial data, like genes associated with a given brain function. A particular challenge when exploring such heterogeneous neurobiological data is the alignment of their spatial reference. Depending on the data acquisition technique, it can be volumetric or region-wise, and different resources are not necessarily in the same reference space. This can lead to time consuming workflows that involve manual aggregation of the data that do not work continuously on different scales.

The entry point for many neuroscience workflows are local brain regions/areas and/or gene expression sites (sites where the gene creates products, such as proteins (Lein et al. 2007)) that are linked to a specific brain function. The functional annotations of such sites are typically the results of neuronal recording, imaging, optogenetics and behavioral neurogenetic studies (e.g. amygdala subnuclei in emotional processing (Haubensak et al. 2010; Kim et al. 2017)). The knowledge of where these local regions/areas- and/or primary expression sites are connected to, is

“This is a post-peer-review, pre-copyedit version of an article published in NeuroInformatics. The final authenticated version is available online at: http://dx.doi.org/10.1007/s12021-019-09428-9

a first step to relate them to a specific brain circuit or a particular function. This information is encoded in so called spatial networks. In these networks, nodes represent regions/areas in the brain, while edges describe their structural (Oh et al. 2014), functional (Betzel and Bassett 2017) or genetic (Richiardi and Altmann 2015) relation/connectivity. Since the size of these networks increases squarely to the number of nodes, these networks can easily grow to hundreds of gigabytes, with billions of edges.

Comparing different types of connectivity is essential for identifying neural circuits. For example, two brain regions can have a high structural connectivity (a connection via neurons) but do not necessarily express the same genes (e.g. a so called ligand-receptor binding (Young and Wang 2004)). Depending on data acquisition techniques, different types of networks are not necessarily available at similar resolution and scale (Betzel and Bassett 2017). Besides being time-consuming, up-sampling networks to higher resolutions requires more storage space, while down-sampling to a lower resolution or even region-level would waive information. When operating on different anatomical scales, i.e. different levels of anatomical parcellation, it is necessary to perform cumulative operations on these networks (e.g. calculate region-wise connectivity from voxel-wise connectivity, aggregate voxel-connectivity of brain areas) to map the networks’ common brain space. In this case large parts of the network need to be loaded and aggregated. The size and complexity of these networks created a need for sophisticated data handling techniques to allow further analyses and exploration (Bassett and Sporns 2017).

Several interactive frameworks for querying connectomic data in neuroscience have been published in recent years. The Allen Brain Institutes’s BrainExplorer as well as its web interface (Oh et al. 2014) can identify pre-computed incoming and outgoing connections of pre-defined locations (injection sites) and anatomical regions in mice. Meso-scale source/target sites are visualized in 3D at voxel level. For quantitative examination this data is ordered by brain region and shown in a list. Although this is an easy-to-use tool for neurobiologists, results cannot be compared directly to other connectivity data or examined with respect to user-generated data. Other tools allow to locally explore the connectomes built by neurons traced on a single EM stack (volumetric electron microscopy) like CATMAID (Saalfeld et al. 2009) and ConnectomeExplorer (Beyer et al. 2013). They are working on a local level of a single network with a fixed scale. This also applies for Sherbseondy et al. (Sherbondy et al. 2005), who used queries on volumes of interest and pre-computed pathways to explore diffusion tensor imaging data, and Tauheed et al. (Tauheed et al. 2013), who developed tree-based spatial management techniques for dense spatial neuron simulations.

The problem of efficiently querying large-scale spatial networks was originally addressed by different domains, particularly on transportation/road networks (Barthelemy 2010). Early approaches in optimizing local queries on road network data were proposed in 1997 by Shekhar and Liu (Shekhar and Liu 1997). In principle, network nodes, and respectively their edges are stored as adjacency list. The list is ordered by a space filling curve, so nodes that are spatially close are stored on the same disk page. This reduces Input/Output (I/O) costs and therefore increases query speed. The data structure was further improved by Papadias et al. (Papadias et al. 2003) and Demir and Aykanat (Demir and Aykanat 2010) with a grid based tree-like hierarchical structure partitioning the spatial domain to efficiently process range queries (perform queries in circular range around a query point) and successor retrieval operations (get all successors of a network node).

Further techniques to speed up network queries can be found in the more general domain of graph computation (Pienta et al. 2015). A common method is the use of advanced caching/paging strategies to hold often accessed parts of a graph in memory (Kyrola, Blelloch, and Guestrin 2012; Han et al. 2013; Roy, Mihailovic, and Zwaenepoel 2013; Chi et al. 2016; Leskovec and Sosič 2016). Other approaches apply memory mapping of large-scale graphs as edge-list files to handle them on the disk programmatically as if they were in the main memory (Lin et al. 2014). This allows for graph processing with billions of edges on consumer level computers and mobile devices (Lin et al. 2014; Lin, Chau, and Kang 2013; Chen et al. 2015). LLAMA (Macko et al. 2015) further uses compressed row storage to harness sparsity. Recent graph computing frameworks such as FlashGraph (Zheng et al. 2015) further facilitate solid-state disks in combination with minimization of I/O operations to perform out-of-memory graph analysis algorithms (Ai et al. 2017).

Despite their universal applicability, the lack of spatial optimization results in inferior performance for *Aggregation Queries*, i.e. aggregated connectivity from, to or between a set of nodes on spatial networks (see *Section Performance Evaluation*). Some of these implementations are tailored to unweighted, binary graphs (Han et al. 2013; Lin et al. 2014; Chi et al. 2016) that are unsuitable to be generalized to perform *Aggregation Queries* on weighted, i.e. non-binary, connectivity data.

"This is a post-peer-review, pre-copyedit version of an article published in NeuroInformatics. The final authenticated version is available online at: http://dx.doi.org/10.1007/s12021-019-09428-9

To our better knowledge, there is currently no tool, which combines those state-of-the art techniques to allow interactive exploration of multimodal, multiresolution neurobiological connectivity on a "big data" level across local-global scales. Thus, from the neuroscientist's perspective, bridging this gap is essential for significant synergies in updating, mining, communicating and sharing brain data.

We meet this demand by proposing a data structure for integration and real time querying of heterogeneous large-scale connectivity matrices at multi-scale voxel and region level by exploiting the hierarchical organization of brain parcellations in combination with spatial indexation.

Region-wise (e.g. resting state functional connectivity) or voxel resolution (structural connectivity, spatial gene expression correlation) connectivity data is aggregated hierarchically, to bridge the gap between different scales and resolutions. The hierarchies are anatomy-driven and can be flexibly generated for different ontologies and their related spatial region annotations. On the lowest level of these hierarchies, high resolution, voxel-wise connectivities with billions of edges (matrices with hundreds of gigabytes) are stored on hard disk in spatially organized indices for high-speed data access. Therefore, aggregated connectivity from, to or between brain areas can be retrieved, from voxel-level to large anatomical brain regions, in an instant.

For direct correlation of different connectivity data at voxel level, we expect the data to reside in the same spatial reference brain space, i.e. registered to the same (multi-resolution) standard brain[1]. However, the dual indexing strategy allows us also to easily integrate and correlate data available only at region level with voxel wise data within the same brain space, but in principle also across brain spaces at region level if the corresponding regions are known. Data from public resources can be easily integrated in our data structure as well as private data generated during experiments in the lab.

We demonstrate the practical significance of this tool by presenting use cases for which we used data provided by large scale brain initiatives. We reproduced recent biological findings by performing data integration and interactive queries on heterogeneous neurobiological data from mice and humans. We created a web-based, interactive local 3D segmentation on visualized data to define volumes of interest (VOI) that can be used to query user-selected connectivity data sets accessible via our data structure. The result is the cumulative voxel-wise connectivity of the selected VOI that is visualized as intensity volume in a 3D rendering. This kind of interaction allows the researcher to relate integrated resources, for example incoming/outgoing connectivity on voxel-level, directly to spatial data like gene expressions.

In general, the proposed data structure allows for handling data of different modalities delivering volumetric and/or connectivity data, which can be used for experimental hypothesis finding. The presented framework is applicable for multilevel functional predictions and extends its relevance across species. Therefore, it is suitable for virtual screening of complex networks, like those linked to psychiatric disorders, and to functionally dissect the corresponding neural correlates in mice.

Materials and Methods

Data

The data relevant for our system can be divided into three types, which in principle can stem from any species or modality:

A hierarchical definition of brain regions and their associated positions on a reference brain. This is basically a hierarchical parcellation of a given standard brain and its related ontology. A hierarchy generally starts with the whole brain divided iteratively into sub-regions, where the lowest level contains the highest resolved regions. These regions can have a dense voxel-level representation (Lein et al. 2007) or a set of coordinates representing biopsy sites (coordinates in the brain from where the gene expression data has been sampled (Hawrylycz et al. 2012)). We exemplarily use the Allen Mouse Brain Atlas (AMBA) ontology with 1288 regions on 5 levels (Lein et al. 2007) on a 132x80x114 voxel space, and Allen Human Brain Atlas (AHBA) ontology with 1840 regions also on 5 levels (Hawrylycz et al. 2012) on 3600 MNI152 coordinate space (representing biopsy sites).

[1] Such a multi resolution reference brain space is e.g. available from the Allen Institute, providing different kinds of data at 100-micron and 200-micron resolution.

"This is a post-peer-review, pre-copyedit version of an article published in NeuroInformatics. The final authenticated version is available online at: http://dx.doi.org/10.1007/s12021-019-09428-9

Connectivity data is given as weighted adjacency matrices. Rows/columns represent the connectivity strength between brain areas on different scales (voxel or region-wise). The weights can be in any range, positive or negative. In the context of the use-cases described in this paper, we used three different types/sets:

1. Structural connectivity: In Ganglberger et al. (Ganglberger et al. 2017) we compiled a voxel-wise structural connectivity matrix that shows the projections (efferent neurons) of ~15% of the brain from *AMBA* and respectively how voxels are structurally connected in a 132x80x114 mouse brain (100 micron resolution, i.e. the side of a voxel has a length of 100-microns). Further details in Supplementary Note 1. The 67500 x 450000 directed connectivity matrix is stored as an uncompressed 91.5 gigabyte *CSV* (comma separated value) file. Weights are normalized to range between 0 and 1.
2. Functional connectivity: Functional connectivity, representing correlation of BOLD fMRI signal shows the functional association of brain regions for specific tasks or resting state. We used a resting state connectome for human (Van Essen et al. 2013) and experimental mouse data, which is only available region-wise (~80 regions). Weights are undirected and represent positive correlation coefficients between 0 and 1.
3. Spatial gene expression correlation networks: Correlated gene expression networks quantify tissue-tissue relationships across genes (Lein et al. 2007; Richiardi and Altmann 2015). Details on matrix creation can be found in Supplementary Note 1. The data consists of a 60000x60000 undirected connectivity matrix for mice, that shows the transcriptional similarity for a specific gene set and 3600x3600 for humans (Hawrylycz et al. 2012). The mouse data has a resolution of 200 microns (67x41x58 voxels mouse brain), and is about 12 gigabyte as uncompressed *CSV* file. The data consists of undirected weights, showing positive correlation coefficients between 0 and 1.

A *Volume of Interest (VOI)* is a spatially related set of coordinates in a reference space. These can be arbitrary selected voxels of a user or a brain region. A *VOI* defines an area in the brain, of which the user would like to know the aggregated source or target connectivity of its individual points.

Managing and Aggregating Hierarchical Connectivity Data

The data access structure we propose is tailored to take advantage of sparseness, anatomical or hierarchical parcellations, and spatial organization of the data, which, to our best knowledge, standard graph managing frameworks such as graph databases are not optimized for.

To allow interactive (real time) exploration of the brain connectivity space, the purpose of the data structure is to retrieve the aggregated source or target connectivity of specific *VOI,* such as anatomical regions or arbitrary user defined areas, on a voxel- or region-level in an instant. These *Aggregation Queries* are executed on connectivity matrices, which we define as weighted directed adjacency matrix

$$C = (c_{ij})_{i=1..|I|,j=1..|J|} \quad , C \in \mathbb{R}^{|\mathbf{I}| \times |\mathbf{J}|}$$

of a graph, where the rows **I** correspond to outgoing-, and the columns **J** to incoming edges of spatial regions, defining the spatial and/or anatomical resolution of the respective connectivity data (which can be voxel level) in the discretized standard brain space $\mathbf{B} = \{\mathbf{p}_x\}_{x=1..n}$, $\mathbf{p}_x \in \mathbb{R}^3$.

Here and in the following, the term *region* refers to a spatial related set of positions in a standard brain space such as certain anatomical brain regions, a group of voxels or a single voxel. Furthermore, we assume the standard brain space to represent the highest occurring resolution of all data to be queried.

Spatial Mapping between Connectivity Matrices and Brain Space

We define the spatial association of the rows, respectively columns of the connectivity matrix **C**, to be a set of ordered disjunct sub-regions (i.e. from anatomical regions or voxel-level), so

$$\mathbf{R}^{\mathbf{ROW}} = \{\mathbf{R}_{\mathbf{1}}^{\mathbf{ROW}}, .., \mathbf{R}_{|\mathbf{I}|}^{\mathbf{ROW}}\}$$

"This is a post-peer-review, pre-copyedit version of an article published in NeuroInformatics. The final authenticated version is available online at: http://dx.doi.org/10.1007/s12021-019-09428-9

$$\mathbf{R}_i^{\mathbf{ROW}} \subseteq \boldsymbol{B},\ \mathbf{R}_i^{\mathbf{ROW}} \cap \mathbf{R}_k^{\mathbf{ROW}} = \emptyset,\ \forall\, i \neq k \wedge\ i,k\ \in \mathbf{I}$$

and respectively column associations $\mathbf{R}^{\mathbf{COL}}$

$$\mathbf{R}^{\mathbf{COL}} = \left\{\mathbf{R}_1^{\mathbf{COL}}, .., \mathbf{R}_{|\mathbf{J}|}^{\mathbf{COL}}\right\}$$

$$\mathbf{R}_j^{\mathbf{COL}} \subseteq \boldsymbol{B},\ \mathbf{R}_j^{\mathbf{COL}} \cap \mathbf{R}_l^{\mathbf{COL}} = \emptyset,\ \forall\, j \neq l\ \wedge\ j,l\ \in \mathbf{J}$$

Note that $\mathbf{C}$ represents voxel-wise connectivity if

$$|\mathbf{R}_i| = 1, \qquad \forall\, \mathbf{R}_i \in\ \mathbf{R}^{\mathbf{ROW/COL}}$$

To directly associate spatial positions in brain space $\mathbf{p}_x \in \mathbf{B}$ with rows and columns of $\mathbf{C}$, i.e. incoming/outgoing connections, we define the following mapping: At first, we map positions in brain space $\mathbf{p}_x$ to the indices of brain regions contained in $\mathbf{R}^{\mathbf{ROW}}$ and $\mathbf{R}^{\mathbf{COL}}$

$$\boldsymbol{\psi}^{\mathbf{ROW}}(\mathbf{p}_x) \coloneqq \begin{cases} i & \text{if } \mathbf{p}_x \in \mathbf{R}_i^{\mathbf{ROW}} \\ \emptyset & \text{if } \mathbf{p}_x \in \mathbf{B} \setminus \mathbf{R}^{\mathbf{ROW}} \end{cases}, \qquad \forall\ \mathbf{p}_x \in \mathbf{B}$$

$$\boldsymbol{\Phi\psi}^{\mathbf{COL}}(\mathbf{p}_x) \coloneqq \begin{cases} j & \text{if } \mathbf{p}_x \in \mathbf{R}_j^{\mathbf{COL}} \\ \emptyset & \text{if } \mathbf{p}_x \in \mathbf{B} \setminus \mathbf{R}^{\mathbf{COL}} \end{cases}, \qquad \forall\, \mathbf{p}_x \in \mathbf{B}$$

This creates non-unique mappings of arbitrary *VOI* in the brain reference space $\mathbf{V} \subseteq \mathbf{B}$ to rows/columns (i.e. a set of positions in the brain space can point to multiple rows or column indices)

$$\boldsymbol{\psi}^{\mathbf{ROW}}(\mathbf{V}) \coloneqq \{i \mid \boldsymbol{\psi}^{\mathbf{ROW}}(\mathbf{p}_x) = i, \qquad \forall\, \mathbf{p}_x \in \mathbf{V}\}$$

$$\boldsymbol{\psi}^{\mathbf{COL}}(\mathbf{V}) \coloneqq \{j \mid \boldsymbol{\psi}^{\mathbf{COL}}(\mathbf{p}_x) = j, \qquad \forall\, \mathbf{p}_x \in \mathbf{V}\}$$

Please note that specific indices in the resulting set might be represented more than once, i.e. if there are m voxels in $\mathbf{V}$ laying in region $\mathbf{R}_i$, then i is m times present. This allows the application of this mapping for aggregation of connectivity.

Vice versa, we map a set of indices to the union of their corresponding regions. As the voxel wise representation of regions in the standard brain space is known, this generates a representation of connectivity at highest voxel resolution independent from the resolution or underlying parcellation of the original connectivity data.

$$\boldsymbol{\psi}^{\mathbf{ROW}^{-1}}(\mathbf{U}) \coloneqq \bigcup_{i \in \mathbf{U}} \mathbf{R}_i^{\mathbf{ROW}}, \forall\, \mathbf{U} \subseteq \mathbf{I}$$

$$\boldsymbol{\psi}^{\mathbf{COL}^{-1}}(\mathbf{W}) \coloneqq \bigcup_{j \in \mathbf{W}} \mathbf{R}_j^{\mathbf{COL}}, \forall\, \mathbf{W} \subseteq \mathbf{J}$$

Therefore, a connection c_{ij} might represents equal connections of several points in brain. This has several advantages (see *Figure 1*):

1. Compare connectivity data defined on different resolutions of the standard brain: $\boldsymbol{\psi}^{\mathbf{ROW}^{-1}}$and $\boldsymbol{\psi}^{\mathbf{COL}^{-1}}$ define the relation of rows respectively columns to voxel at a certain resolution. Since the overlap of regions with standard brain space is known, this enables a comparison of connectivity matrices in respect to different brain parcellations and/or different resolutions. If the resolution is smaller than the reference space, this mapping would represent up-sampling (see *Figure 1A*).
2. Map region wise connectivity to voxel level: Nodes of a connectivity matrix can also represent (anatomical) brain regions to store region-wise connectivity data. Using $\boldsymbol{\psi}^{\mathbf{ROW}^{-1}}$and $\boldsymbol{\psi}^{\mathbf{COL}^{-1}}$ allow a retrieval of the data in voxel-wise brain space and therefore also allow the comparison of connectivity with respect to different brain parcellations (see *Figure 1B*).
3. Build caches: This technique can also be used to store precomputed data, such as connectivity of brain regions (from voxel level data) or pyramids representations with lower resolution (like an image pyramid). Although this increases the required storage, it improves scalability (see *Figure 1C*).

“This is a post-peer-review, pre-copyedit version of an article published in NeuroInformatics. The final authenticated version is available online at: http://dx.doi.org/10.1007/s12021-019-09428-9

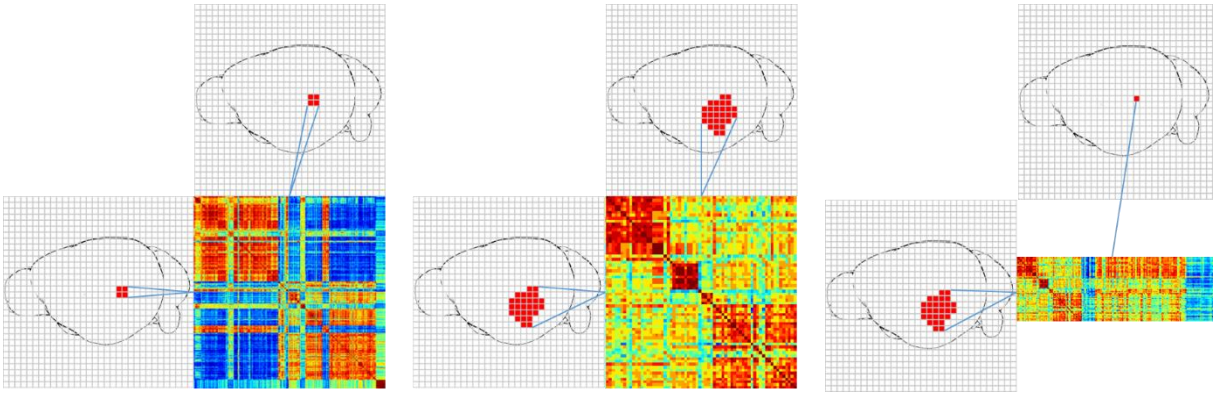


Figure 1 A: Connectivity matrix in a 4 times lower resolution than the reference brain space. Therefore every row/column is associated with 4 voxels. ***B***: Region-wise connectivity matrix. Every row/column is associated with voxels that form brain regions. ***C***: Region cache. Preprocessed aggregated outgoing connectivity for brain regions on voxel level.

A Dual Data Stucture Strategy for Aggregation Queries

Aggregation Queries are defined as follows. Let $\mathbf{V} \subseteq \mathbf{B}$ be a *VOI*. The result of a target aggregation query is the cumulated outgoing connectivity for every position in space $\mathbf{B}$

$$\boldsymbol{\tau}(\mathbf{V}) = \left(\sum_{i \in \boldsymbol{\psi}^{\mathbf{ROW}}(\mathbf{V})} \mathbf{c}_{\mathrm{i},\boldsymbol{\psi}^{\mathbf{COL}}(\mathbf{p_x})} \right)_{\mathbf{p_x} \in \mathbf{B}}$$

and the result of a source aggregation query the cumulated incoming connectivity for every row

$$\boldsymbol{\varsigma}(\mathbf{V}) = \left(\sum_{j \in \boldsymbol{\psi}^{\mathbf{COL}}(\mathbf{V})} \mathbf{c}_{\boldsymbol{\psi}^{\mathbf{ROW}}(\mathbf{p_x}),\mathrm{j}} \right)_{\mathbf{p_x} \in \mathbf{B}}$$

We are proposing a dual strategy unifying two complementary data structures to efficiently realize *Aggregation Queries*. The *Connectivity Storage* handles the data access for the *Aggregation Queries*, and the *Region-Wise Connectivity* in a Graph-Database manages queries on (anatomical brain-)region level. *Figure 2* gives an overview of the overall system. Incorporating a connectivity matrix into our data structure begins with a preprocessing, that harnesses spatial-organization of the data (*Figure 2 (1)*) and uses row-compression to minimize disk-space (and therefore reading-time for queries) (*Figure 2 (2)*) to create a *Connectivity Storage File*. Region-wise connectivity of a hierarchical anatomical brain-region parcellation is precomputed and stored in a graph-database *Figure 2 (3)*). To further improve query performance, *Connectivity Cache Files* are created, that store pre-computed connectivity for faster data access (*Figure 2 (4)*). Voxel-wise connectivity can then be queried from cache files and *Connectivity Storage Files* (*Figure 2 (5)*), region-wise connectivity form the graph-database (*Figure 2 (6)*). Preprocessing (*Figure 2 (1,2,3)*) is further described in the following subsections (*Connectivity Storage, Region-wise Connectivity Database)*, and cache-creation as well as querying (*Figure 2 (4,5,6)*) in subsection *Implementation.*

“This is a post-peer-review, pre-copyedit version of an article published in NeuroInformatics. The final authenticated version is available online at: http://dx.doi.org/10.1007/s12021-019-09428-9

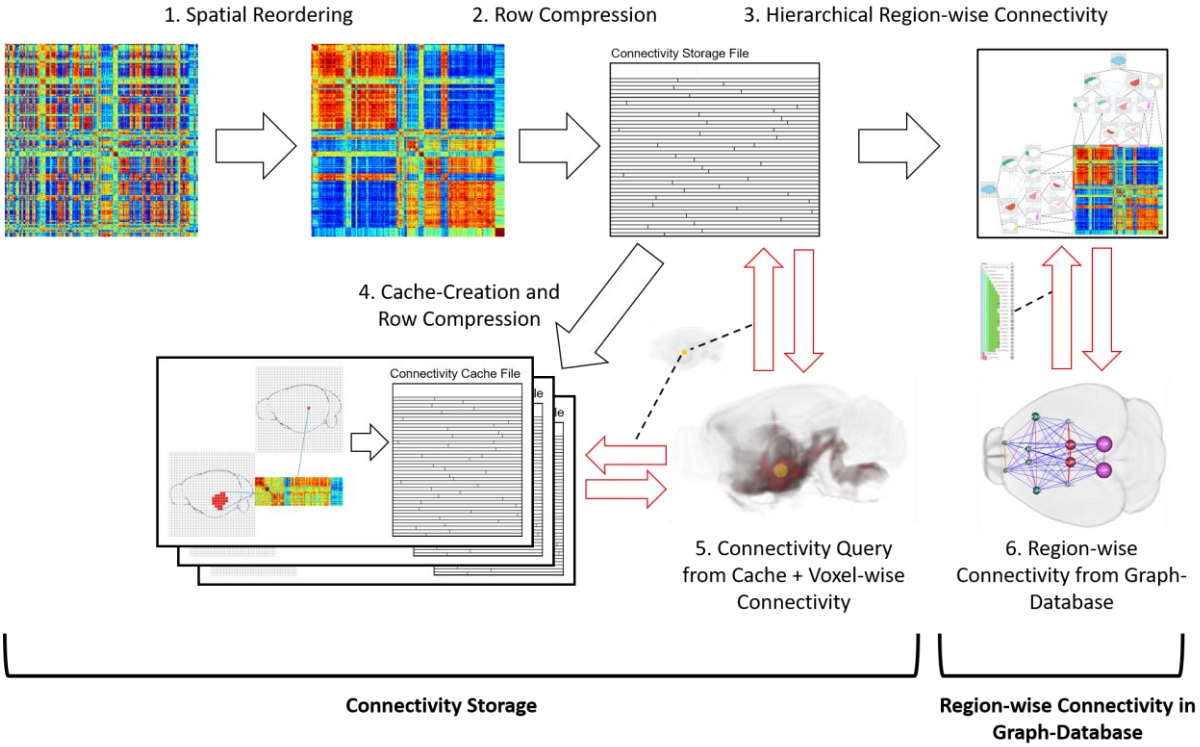


***Figure 2*:** Overview of Connectivity Storage and the Region-wise Connectivity in the Graph-Database. Black arrows: Preprocessing of the data. **1.** Spatial Reordering of a (voxel-wise) connectivity matrix with a space filling curve. **2.** Row-wise compression of spatially-ordered connectivity matrix. 3. Generation of hierarchical region-wise connectivity and storage in graph-database. **4.** Cache creation (preprocessed voxel-wise connectivity for predefined regions), and storage with row compression. Red arrows: **5.** Querying a *VOI* (yellow circle) on *Connectivity Cache Files*, then on *Connectivity Storage File*, resulting in aggregated connectivity (red). **6.** Querying connectivity between preselected brain regions (from a hierarchical parcellation), resulting in a region-wise connectivity graph.

Connectivity Storage: Since *Aggregation Queries* involve the reading and aggregation of whole rows or columns of connectivity matrices, we use a row-wise storage scheme. Allthough edge lists are popular for many graph management tools (Lin et al. 2014), which store connections in a <source node, target node, value> combination, they create a significant storage overhead for dense connectivity matrices.

Reducing data size allows higher query speed, since fewer data needs to be read. Therefore we apply a row-wise compression, that exploits potential sparseness of the data. First, the rows and columns of **C** are ordered by a space filling curve (Hilbert 1891) to preserve locality. The reordering causes sparse/dense areas to cluster within each row/column, since the connectivity of a region/voxel is not randomly distributed over the brain, but spatialy related. Then, a compressed row starts with the column index of the first non-zero value (*NZV*), the amount of *NZV* to follow, and the following *NZV*s. This is repeated similarly with the column index of the next *NZV* until the end of the row is reached. To identify each row in the file, an additional mapping

$$\Omega(i) \coloneqq f \quad , \qquad \forall\, f \in \mathbf{F}, i \in \mathbf{I}$$

needs to be created, depicting the beginning of each row to their position f in the file **F**. A connection c_{ij} can be identified by going to the corresponding position of the i-th row f, reading the j-th value from the row-wise compression. *Figure 3* illustrates this process.

“This is a post-peer-review, pre-copyedit version of an article published in NeuroInformatics. The final authenticated version is available online at: http://dx.doi.org/10.1007/s12021-019-09428-9

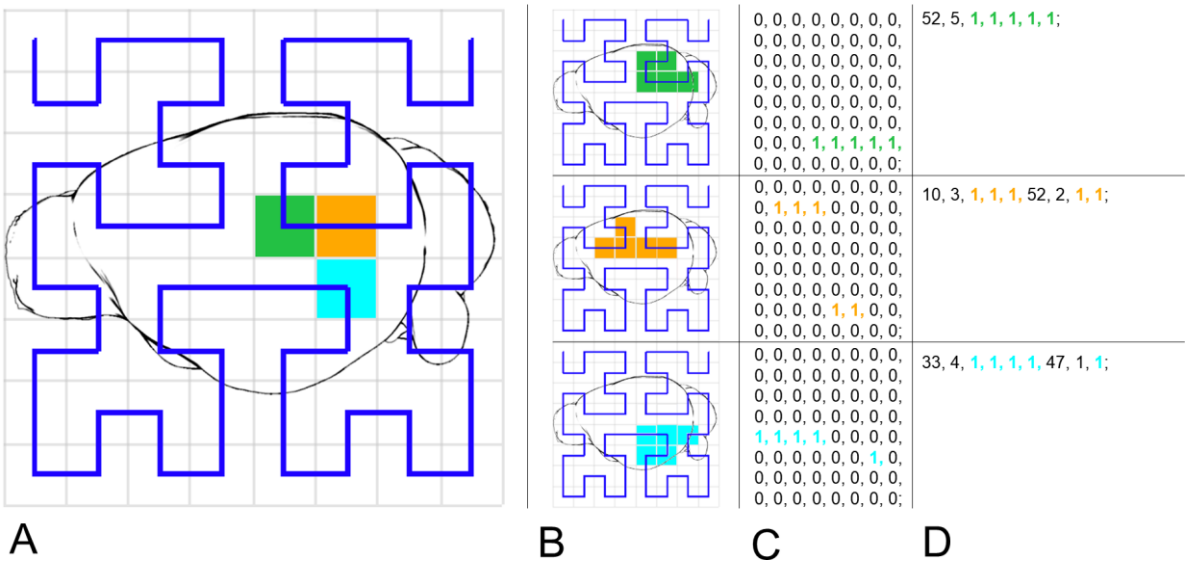


Figure 3 A: A mouse brain with overlayed Hilbert curve (blue), mapping the space to a one-dimensional space ***B****:* The outgoing connectivity of the voxels corresponding to the colored voxels in *A.* These would represent 3 rows in a connectivity matrix. ***C****:* The connectivity of B along the Hilbert curve (for simplicity in this example, connectivity is either 0 or 1). ***D:*** Row-wise compression of *C.* The compression can be read this way: On the 52nd position, 5 *NZVs* are following (green). On the 10th position, 3, and on the 52th position 2 *NZVs* are following (orange). On the 33rd position, 4 *NZVs* and on the 47th position, there is 1 *NZV* following (cyan).

Other compression methods would also reduce the data size, but would not allow to directly access single rows without decompression of the whole file or significant parts of the file (Barrett et al. 1994).

For every connectivity matrix, we create a separate *Connectivity Storage File*, consisting three indices as header (*FILE*: $\boldsymbol{\Omega}(\mathbf{i})$, *ROW:* $\boldsymbol{\psi}^{\mathbf{ROW}}(\mathbf{V}), COLUMN: \boldsymbol{\psi}^{\mathbf{COL}}(\mathbf{V})$ followed by compressed rows (*Figure 4B*). Even after compression, the compressed file does not necessarily fit into memory, especially when one wants to query on multiple connectivity matrices. We use memory mapping (*MMap*) to map the file into virtual address space. This allows to programmatically access rows if they were in the main memory, without overhead of system calls. Furthermore, the OS employs paging strategies, such as read-ahead paging. When performing a query for outgoing connections, the rows can be read in the order of their position in the file, and directly benefit from read-ahead paging of the operating system to reach near-sequential reading speed. This additionally exploits the spatial organization of the data, that has been created with the ordering by space filling curve (see *Figure 4A*). Multiple connectivity matrices can then be queried sequentially without loading the whole matrices into memory. Note that a connectivity matrix of a directed graph needs an additional transposed *Connectivity Storage File* to query incoming connectivities (for undirected graphs, outgoing and incoming connections are equal due to symmetry).

“This is a post-peer-review, pre-copyedit version of an article published in NeuroInformatics. The final authenticated version is available online at: http://dx.doi.org/10.1007/s12021-019-09428-9

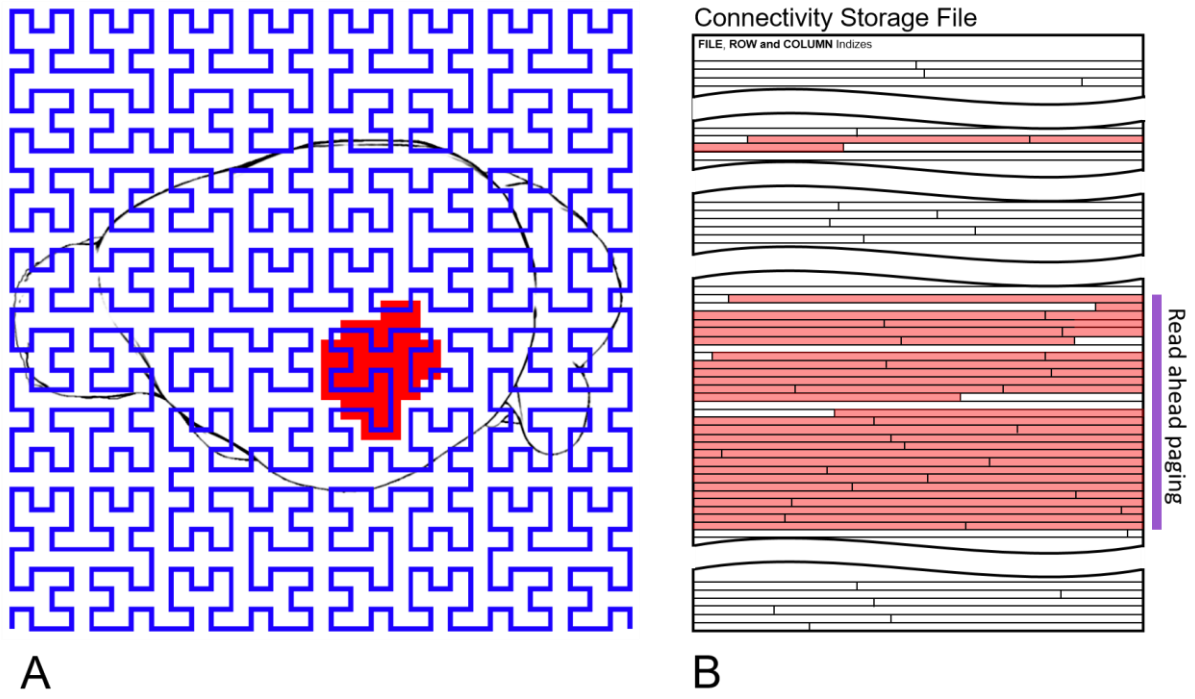


Figure 4 A: Brain Space overlayed with Hilbert curve (blue) and a *VOI* (red) that is queried for outgoing connectivity, ***B:*** Connectivity matrix file (rows ordered by Hilbert curve). Red-blocks represent rows that are read in order to get outgoing connectivity of *VOI* shown in *A*. Blocks can be read sequentially. Purple rows benefit by read ahead paging.

Region-Wise Connectivity Database: On higher levels, the (anatomical brain-)region level, the aggregated connectivity of a region consists of the connectivity of its subregions (and on the lowest level voxel-wise connectivity). When looking at brain wide region-wise graphs, it is not feasible to read the entire *Connectivity Storage* and compute the connectivity hierarchically at runtime. This would be too resource consuming for real-time computation. Instead, we compute it once when the *Connectivity Storage* is created. The resulting region-wise hierarchical connectivity is stored in a graph-database. The region-wise connectivity is computed recursively bottom up: First, the lowest level regions are aggregated from the *Connectivity Storage*, then the regions above are aggregated by their levels below until the top of the hierarchy. We further compute the connectivity between the levels in a similar way. Therefore, it is not necessary to compute any region-level connectivity at runtime (*Figure 5*).

“This is a post-peer-review, pre-copyedit version of an article published in NeuroInformatics. The final authenticated version is available online at: http://dx.doi.org/10.1007/s12021-019-09428-9

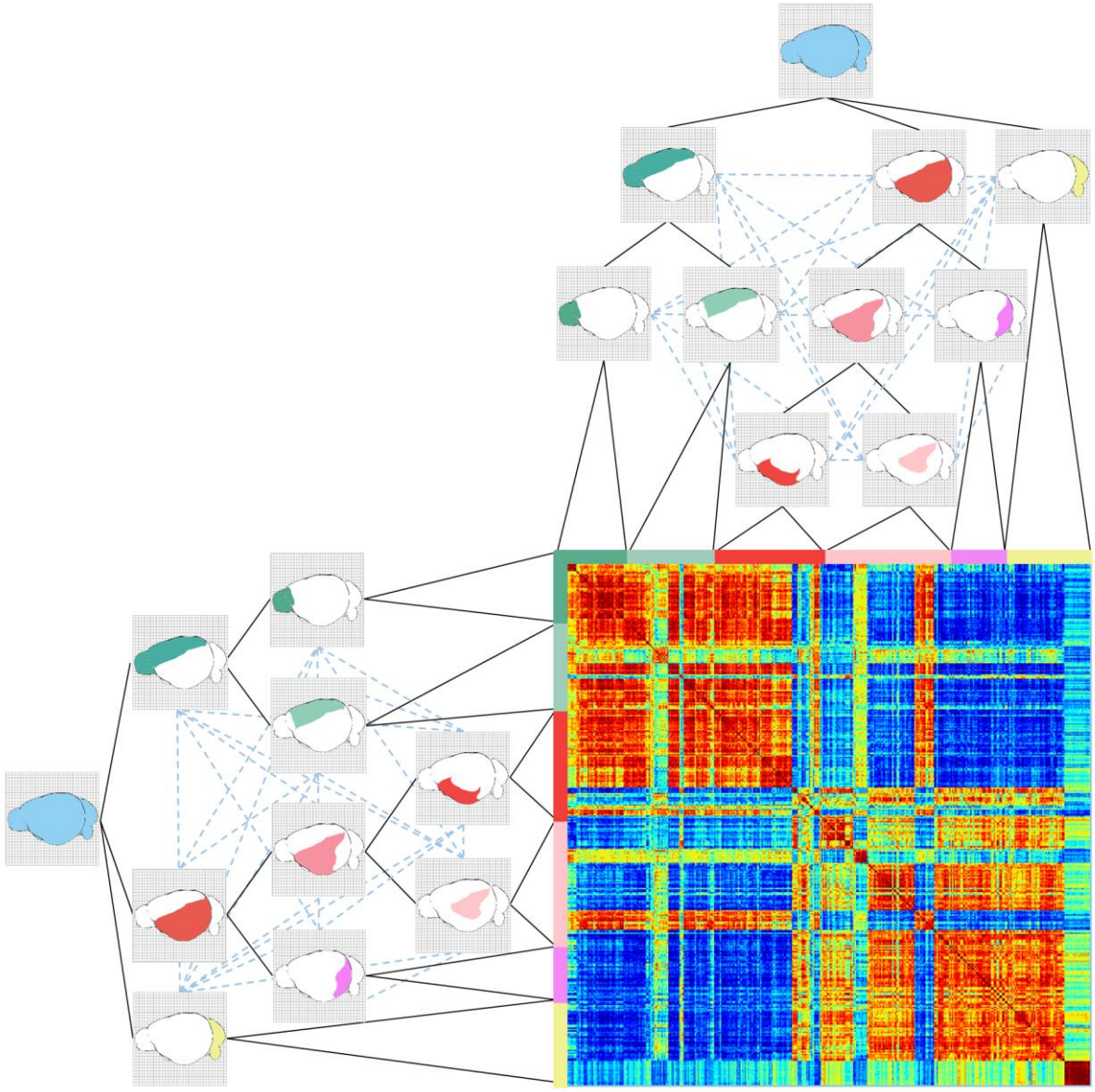


Figure 5: Scheme of the data structure: The *Connectivity Storage* stores the connectivity on the lowest level (voxel-wise connectivity). Region-wise connectivity (dotted blue lines) is aggregated from the *Connectivity Storage* hierarchically.

Implementation: As central access point for the data, we created a REST API in GO (golang). It provides calls for importing data, creating caches as wells as *Aggregation Queries*. These are executed on the *Connectivity Storage*, which was implemented in C++ for memory and performance optimization. Connections are stored in a 4-byte floating point format, which supports a range of values $\pm 1.18 \times 10^{-38}$ to $\pm 3.4 \times 10^{38}$, with single precision (about 7 decimal digits). We choose this as trade-off to storage space, since higher precision would also cause higher reading times. *ROW* and *COL* indices have a 4-byte unsigned integer format. Therefore the maximum amount of edges is limited by 4294967295 x 4294967295 (= 1.84467×10^{19}). The *FILE* index associates rows the with 8-byte unsigned integers to file positions, limiting the file size similarly to 1.84467×10^{19} connections or 64 petabyte.

We implemented two types of *Connectivity Caches* to increase performance: A region-cache, that stores the aggregated voxel-level connectivity of lowest-level of the hierarchical brain-region parcellation, and factor h low-resolution versions ($h \in \mathbb{N}, h \leq |\mathbf{I}|$) of the *Connectivity Storage*, which cumulates the connectivity of h voxels along the Hilbert curve (basically every h rows of the *Connectivity Storage* being aggregated).

“This is a post-peer-review, pre-copyedit version of an article published in NeuroInformatics. The final authenticated version is available online at: http://dx.doi.org/10.1007/s12021-019-09428-9

When executing an *Aggregation Query* for a *VOI*, $\boldsymbol{V} \subseteq \mathbf{B}$, the *Connectivity Cache Files* will be accessed first to check if the *VOI* contains cached regions $\mathbf{R}^{\mathbf{ROW\ (CACHE)}}$ defined in the *ROW* index of the cache. The connectivity of a region $\mathbf{R}_{\boldsymbol{c}} \in \mathbf{R}^{\mathbf{ROW\ (CACHE)}}$ will be added to the results from the cache, if $\mathbf{R}_{\boldsymbol{c}} \subseteq \mathbf{V}$, i.e. all spatial positions of a region are contained within the *VOI*. Before the *Connectivity Storage* will be accessed, all *Connectivity Cache Files* will be queried until no further region in the cache can be found. Only after this, the remaining brain space positions of the VOI will be queried from the *Connectivity Storage*, hence, the total number of row-reads is minimized.

The anatomical hierarchy is represented in *OrientDB* (Garulli 2010), a graph database that can be used to store further region information, such as masks, 3D models or links to online repositories. Region-wise connectivities within those hierarchies consist of 1000-2000 regions with a maximum of 4 million edges. When querying such comparatively small graphs, the performance differences of standard graph databases to the *Connectivity Storage* is neglectable. Therefore, we store them in *OrientDB*, where it is directly linked to the brain regions.

To access the API, we created a web-component that allows visual queries that are based on selections of *VOI* directly in 2D slice views, visualized simultaneously in a 3D volume rendering. Via a spherical brush tool, a user-defined area can be marked. *Figure 6 A* shows for example a gene-expression volume, where the spherical area is drawn on voxel with high gene-expression. After selection, *Aggregation Queries* can be used to link connectivity data with volume data. The selected area (*Figure 6 B*), is used as input for an *Aggregation Query* on the API. The API retrieves the connectivity from the *Connectivity Storage* to all voxels that are either *targets* or *sources* of the selection, and the web component will instantly render the connectivity as volume. This represents the cumulative connectivity to (target) or from (source) the selected area (*Figure 6 C*). Furthermore, the connectivity can be quantified in *Connectivity Profiles*, which shows the cumulated connectivity of the *VOI* to preselected (brain) regions (*Figure 6 D).*

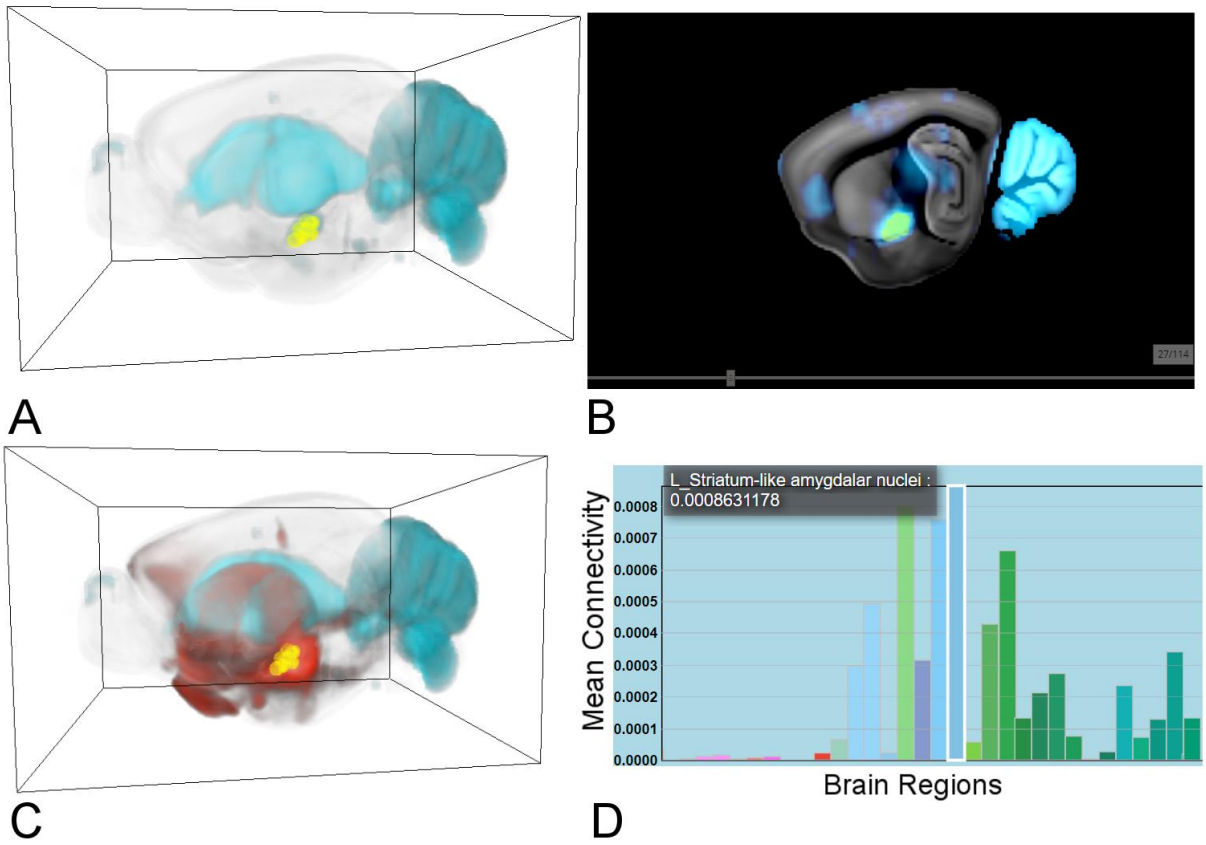


Figure 6 A*:* gene expression (cyan) with brush selection (yellow) of *VOI* in 3D ***B****:* Selection has been performed on 2D slice views, ***C***: Accumulated target connectivity (red) of *VOI* in 3D ***D*****:** *Connectivity Profile* of the target query, showing the mean connectivity to each brain region.

“This is a post-peer-review, pre-copyedit version of an article published in NeuroInformatics. The final authenticated version is available online at: http://dx.doi.org/10.1007/s12021-019-09428-9

Results

To assess the efficiency and effectiveness of the data structure in context of its practical application, we performed a quantitative and qualitative evaluation on real world data that was introduced above in *Section Data*. We quantified the effect of the data structure's parameters (row-compression, spatial ordering, caches) on query performance and compared these results with two state-of-the-art graph engines (*Section Performance Evaluation)*. We further performed two Case-Studies that we designed with domain experts in order to demonstrate the relevance of the data structure for neuroscientific research *(Section Case Study 1 and 2*).

Performance Evaluation

To verify the data structure's applicability for real-time *Aggregation Queries*, we created test queries on three voxel-level connectivities which were introduced in *Section Data*. We used one directed structural connectivity matrix *SC*, resulting in two *Connectivity Storage Files* for targets and source queries, and two undirected spatial gene expression correlation networks *CS1* and *CS2* (which are further used in Case Study 1 and 2), creating one *Connectivity Storage File* each (because they are undirected).

Creating the two *Connectivity Storage Files* for SC (91GB CSV file) took 32 minutes in total (19 for the first, 13 min for the transposed) while CS1 (12 GB CSV file) and CS2 (13 GB CSV file) took about 3 min each on an SSD with our REST API. Therefore, the file creation takes approximately 21 sec/GB for directed, and 13 sec/GB for undirected matrices. Generating the *Region-Wise Connectivity Database* lasted less than 10 minutes for each *Connectivity Storage* which depends on the I/O performance of the *OrientDB.*

In cooperation with domain experts, we defined 10 queries in our web-component (user-queries), which are shown in Supplementary Note 2. The *VOI* of these queries range from 0.2% to 10% of the mouse brain space. In addition, we selected 10 distinct anatomical brain regions to act as *VOI* (region-queries) with sizes ranging from 0.2% to 4% (see Supplementary Note 3). To evaluate queries on a bigger scale, we further created 100 random queries by using randomly placed spheres with random radii as *VOI*s. The sizes of these range from 0.2% to 5%, because it was not possible to place larger spheres within the mouse brain space.

We used these queries to assess the effects of individual components of the *Connectivity Storage*, such as row-compression, the spatial-ordering of rows/columns and *Connectivity Caches.* To demonstrate the data structure's relevance for performing *Aggregation* Queries, we compared the results to the state-of-the art tools FlashGraph (Zheng et al. 2015) and GraphChi (Kyrola, Blelloch, and Guestrin 2012). We did not evaluate the performance of the *Region-wise Connectivity Database* in the *OrientDB* specifically, since retrieving a connection between two regions only involves accessing a single database entry (<10ms), in comparison to aggregating mega- to gigabytes of data from the *Connectivity Storage*.

Performance has been evaluated on an Ubuntu 16.10 64-bit machine with Intel Core i7-4470 CPU, 32 GB RAM and a 1 Terabyte SSD with a sequential read-speed of 520 MB/sec. Test result on an HDD with 120 MB/sec sequential read-speed can be found in Supplementary Note 4.

Effect of compressed row-storage on data size: For 3 connectivity matrices (*SC*, *CS1* and *CS2*), we created 4 *Connectivity Storage Files* (2 for *SC* and 1 for each *CS1 and CS2*). *Figure 7A* shows that *Connectivity Storage Files* with compression reduces the initial file size of *SC* by half, even if one is using random ordering of rows/columns. Spatial ordering by a Hilbert-curve further improves file size by reducing it by half. The effect is smaller for *CS1* and *CS2*, since they are not as sparse as *SC* (i.e. they contain not as many zeroes).

Effect of spatial-ordering on query speed: We executed the user-, random-, and region-queries on *SC*, *CS1* and *CS2* for their sources and target connectivity. *Figure 7C* shows the mean query time and their standard error bars on the connectivity matrices for different query types. Not, that the spatial ordering along a Hilbert curve greatly reduces query-time compared to random-ordering, especially for the bigger *SC* matrix (from up to 20 seconds to <2 seconds). This is due to read-ahead-paging, which benefits from sequential reading. Note that the mean query time for different query types depends on the size of their *VOI*. Hence, region-queries are faster than user- or random-queries simply because they involve reading fewer data (detailed query sizes see *Figure 7B).*

“This is a post-peer-review, pre-copyedit version of an article published in NeuroInformatics. The final authenticated version is available online at: http://dx.doi.org/10.1007/s12021-019-09428-9

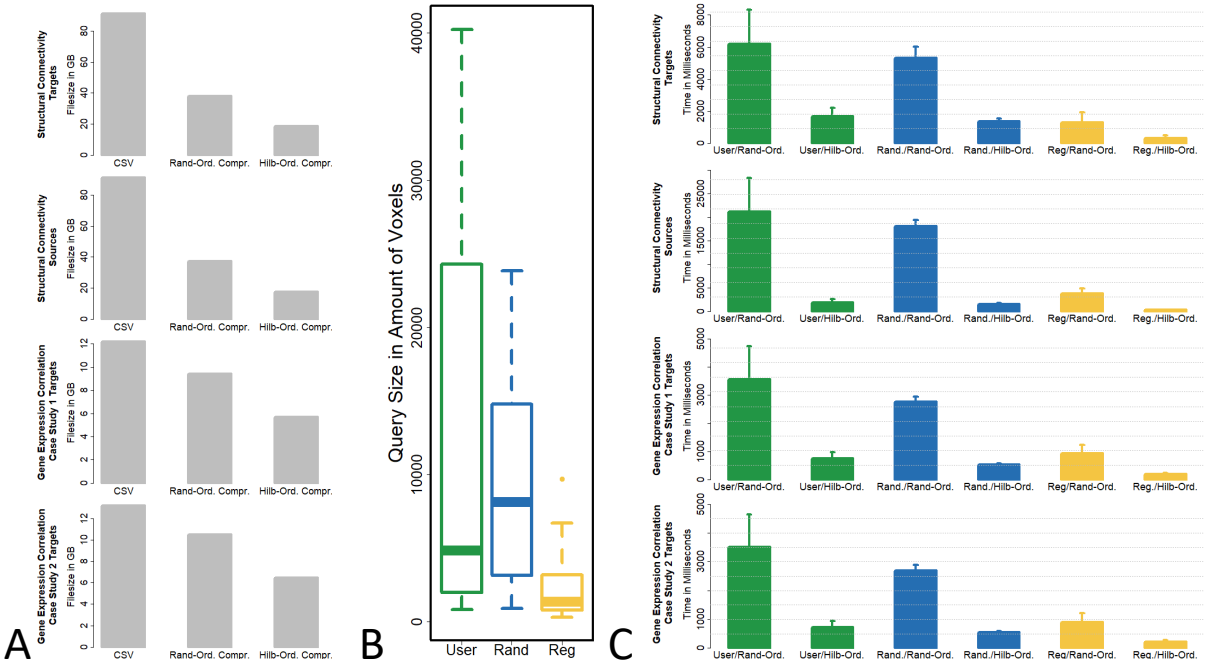


Figure 7 A: Effect of compressed row storage on the data size of different connectivity matrices. Bars indicate the size of the original CSV, the *Connectivity Storage* file without compression, random-ordering with compression and Hilbert-ordering with compression. **B:** Boxplot of the *VOI* size as amount of voxels (i.e. the query size of 10 user-defined *VOI* queries (green), 100 random *VOI* queries (blue) and 10 region *VOI* queries (yellow) **C:** Effect of spatial-ordering on query-speed on different connectivity matrices. Bars show the mean query-time with standard error of 10 user-defined *VOI* queries (green), 100 random *VOI* queries (blue) and 10 region *VOI* queries (yellow), for Hilbert-Ordering and Random-Ordering.

Effect of *Connectivity Caches* on query speed: As described in *Section Implementation*, we created a region *Connectivity Cache* of the lowest level of the hierarchical brain-region parcellation, and factor h low-resolution *Connectivity Caches*, where every h rows of the *Connectivity Storage* are aggregated, for h =10 and h =100. *Figure 8* shows the mean query time and its standard error for different cache combinations. One can see that for high resolution *Connectivity Matrices* such as *SC*, h-factor caches can save up to half of the query time, while region queries especially benefit from the region-caches. For lower resolutions (*CS1* and *CS2*), this effect of h-factor caches is not as strong. The reason is that connectivity retrieved from *Connectivity Caches* leaves “holes” in the *VOI* of the query, hence, the remaining rows that need to be read from the *Connectivity Storage File* are fragmented. This reduces the overall read speed, for it relies on read-ahead paging (the effect of sequentially reading spatially close rows has been shown in *Figure 7C*). *Figure 9* depicts this in further detail: In the left column, one can observe that the query time depends on the query size, and that query time benefits increasingly from *Connectivity Caches* for larger query sizes (i.e. more data to read leads to higher chances of read-ahead paging). The right column shows read-speed on *Connectivity Storage Files* vs query size and therefore the effect of reads from the *Connectivity Caches* and the resulting fragmentation. The lower read speed after cache reads is a direct cause of the higher fragmentation rate (i.e. many reads from 10-factor caches lead to more “holes” in the *VOI* than a few reads from 100-factor caches).

“This is a post-peer-review, pre-copyedit version of an article published in NeuroInformatics. The final authenticated version is available online at: http://dx.doi.org/10.1007/s12021-019-09428-9

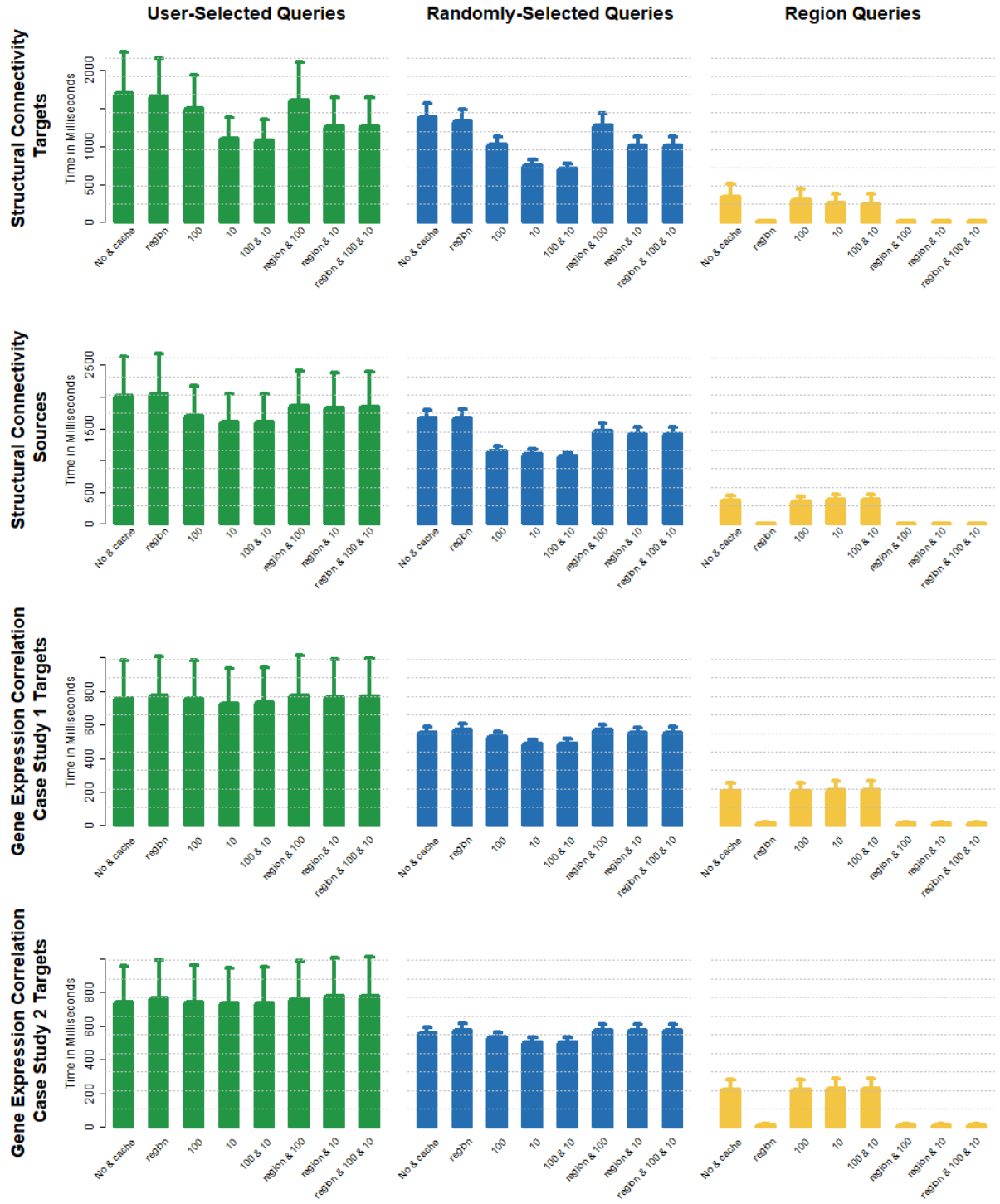


Figure 8: Effect of *Connectivity Cache* on query-speed on different connectivity matrices. Bars show the mean query-time with standard error of 10 user-defined VOI queries (green), 100 random *VOI* queries (blue) and 10 region *VOI* queries (yellow), for different types of caches and their combination.

“This is a post-peer-review, pre-copyedit version of an article published in NeuroInformatics. The final authenticated version is available online at: http://dx.doi.org/10.1007/s12021-019-09428-9

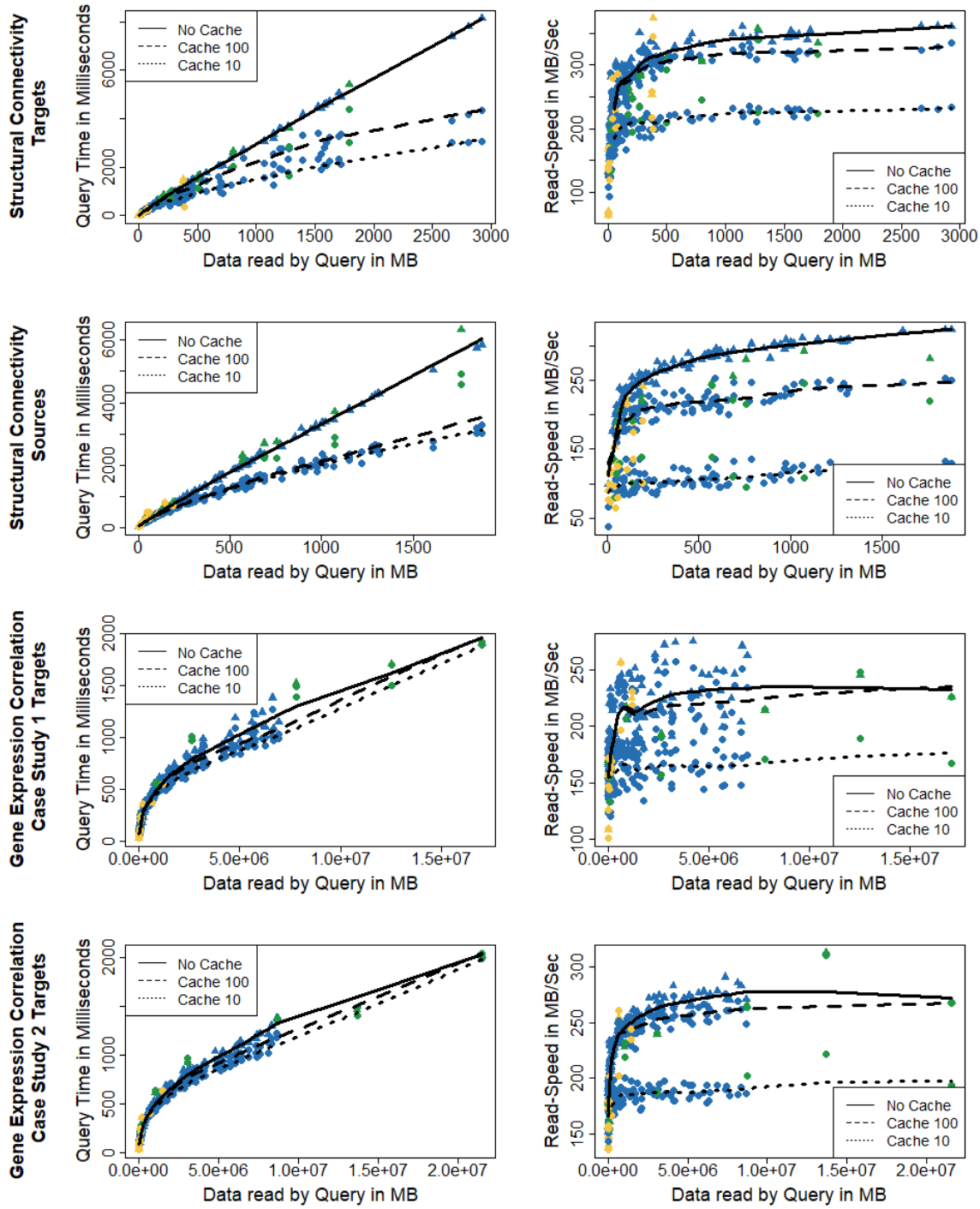


Figure 9: Relation of query-time and read-speed on query size for different *Connectivity Matrices*, *Connectivity Caches* and query types. The left column depicts the query time vs query size for queries executed with cache (●) and without (▲), while the color depicts the query type (green=user query, blue=random query and yellow=region query). The right column depicts read speed on the *Connectivity Storage Files* vs query size, similarly encoded. LOWESS regression lines are added to see the overall trend for different cache sizes.

Comparison to state-of-the-art tools: We compared our method to the state-of-the-art graph engines FlashGraph (Zheng et al. 2015) and GraphChi (Kyrola, Blelloch, and Guestrin 2012). Both tools are capable of computing graph algorithms (page-rank, breath-first-search etc.) on graphs with billions of edges on consumer level machines (i.e. without hundreds of gigabytes RAM). They achieve this by utilizing data access mechanisms that are able to load data from hard-drive on demand, instead of holding the whole graph in memory. GraphChi’s approach is

"This is a post-peer-review, pre-copyedit version of an article published in NeuroInformatics. The final authenticated version is available online at: http://dx.doi.org/10.1007/s12021-019-09428-9

splitting the data into small parts (so called shards), and loading them on demand, while FlashGraph uses optimized I/O requests for SSDs. Therefore, these methods benefit from graph queries that do not involve whole graphs respectively, do not need to load entire connectivity matrices, such as *Aggregation Queries*. To compare their performance to the *Connectivity Storage*, we have implemented *Aggregation Queries* for both (see Supplementary Note 5 for details) and created edge-lists (in their common input data format <source node, target node, value>) of our connectivity matrices. Further, we have ordered the node indices spatially (according to a Hilbert curve) to test them under equal conditions. *Figure 10* shows that even with Hilbert ordering, FlashGraph and GraphChi do not perform as fast as our method. While on smaller graphs (CS1 and CS2), the *Connectivity Storage* is still faster than FlashGraph by a factor of 2-3, this effect is even stronger for larger matrices (SC1) with a factor of 6. Overall, our method performs more than 5 times faster than FlashGraph, and 160 times faster than GraphChi. One has to note, that these tools were developed for performing various graph analysis methods, thus, they are probably not optimized for *Aggregation Queries.* Especially GraphChi is more suited for analyzing whole graphs, while *Aggregation Queries* only require loading of subgraphs.

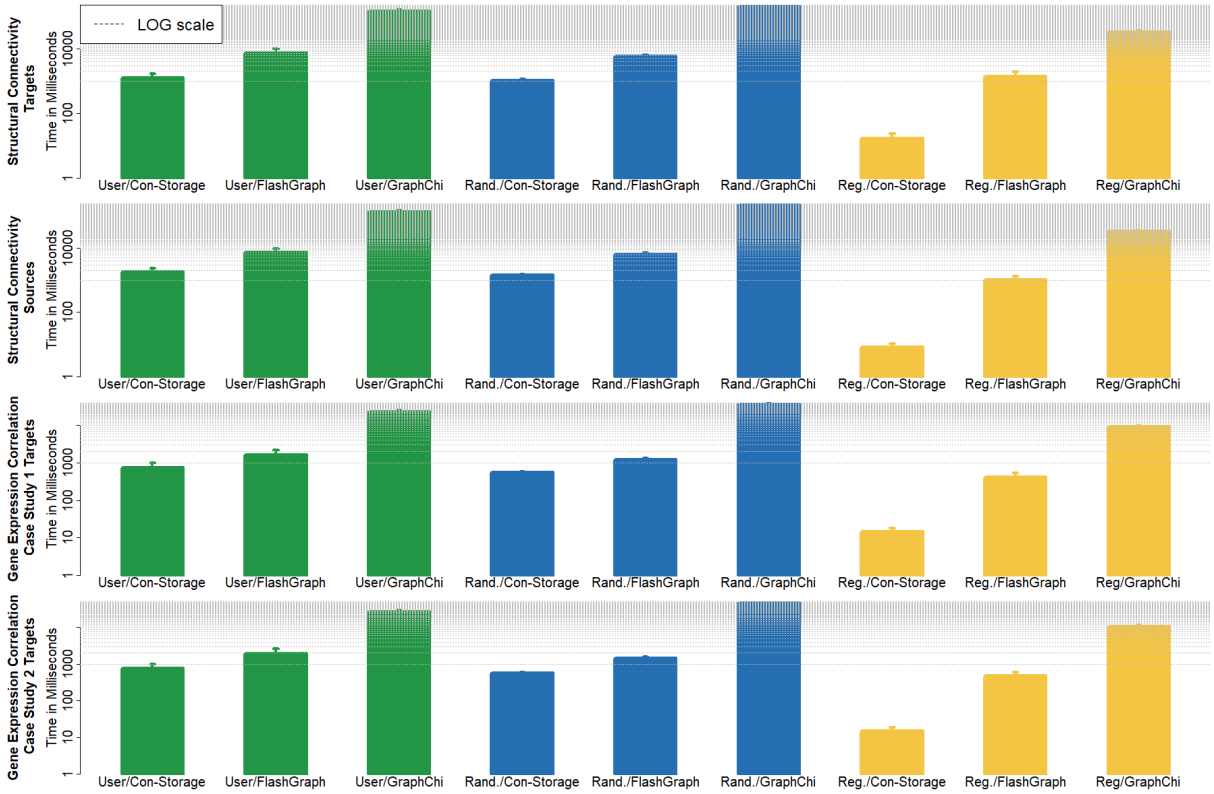


Figure 10: Comparison of query speed with state-of-the-art tools. Bars show the mean query-time with standard error of 10 user-defined VOI queries (green), 100 random *VOI* queries (blue) and 10 region *VOI* queries (yellow), for the Connectivity Storage, FlashGraph and GraphChi. The bars are log scaled, indicated by equidistant grey dotted lines (distance between two lines represent 1 second).

Example video for real-time performance: For further demonstration, Supplementary Video 1 shows a target query on the structural connectivity matrix (similar to *Figure 6)* performed in real-time.

Case Study 1: Exploring different types of connectivity emerging from a brain area of interest

This case study has been chosen for its particular application in circuit dissection. Recent advances in circuit neuroscience (e.g. neuro- and behavioral genetics, optogenetics, imaging) identified gene sets underlying specific behavioral function. Hence, we mapped such function-related network context on a genetically well dissected microcircuitry (Radke 2009). To illustrate this case, we focused on the central amygdala (CEA), an amygdala subnucleus and hotspot expressing several functionally related genes, whose role in fear behavior is a heavily researched topic in the neuroscience community.

“This is a post-peer-review, pre-copyedit version of an article published in NeuroInformatics. The final authenticated version is available online at: http://dx.doi.org/10.1007/s12021-019-09428-9

The connectivity data used for this case study consists of directed structural connectivity (*Data Set 1)* and undirected spatial gene expression correlation (*Data Set 3*). Hence, this case demonstrates the exploration of connectivities of different type and different resolution.

The entry point for our experts is a subset of these genes consisting of Prkcd (EntrezID: 18753), Sst (20604), Crh (12918), Dyn (18610) and Penk (18619) that have been known to regulate fear responses (Haubensak et al. 2010). We examined the gene expression density of these genes in 3D and 2D slice views for areas of high co-expression (where multiple genes are expressed). An image overlap of Prkdc, Crh and Dyn revealed an enclosed area (*Figure 11 A, red arrow)* that is selected by using a brushing tool allowing the user to interactively mark *VOI*s on 2D slice views of the brain space. We further overlaid the outlines of CEA so the selected area is this brain region indeed (*Figure 11 B, red arrow).*

After a target query on the structural connectivity (*Data Set 1)* matrix (*Figure 11 C)*, which is performed in less than a second, particularly strong connected areas are visualized and identified by *Connectivity Profiles*. It highlights, that among other regions, the bed nucleus of the stria terminalis (BNST) has a strong connection to the central amygdala (CEA) (*Figure 11*, red arrow). It is important to note, that this confirms known structural anatomy from literature (Radke 2009). Interestingly, the BNST is functionally related to CEA. While CEA causes brief phasic fear responses, BNST shows more long-lasting tonic anxiety-like states. Thus, this approach recaptures a functional CEA-BNST circuit module for fear.

To further verify the query’s result quantitatively, we compared the outgoing connectivity to known region-level structural connectivity of CEA. Therefore we used the normalized projection strength of 469 sites (positions in the brain) to 590 brain regions provided by Oh et al (Oh et al. 2014) (Figure 3/Supplementary Table 2), i.e. the cumulated outgoing strength of projection neurons. Out of these 469 sites, we choose the five that lie within CEA since there is a high overlap (*Figure 11 B, red arrow*) between the query’s VOI and CEA.
Figure 12: Correlation between mean outgoing connectivity of the query’s VOI (Figure 11 B) and the normalized projection strength of sites within CEA according to Oh et. al Figure 3/Supplementary Table 2 (Oh et al. 2014). depicts the rank correlation of the query result to the five sites chosen by us, as well as the mean connectivity thereof. A correlation of 0.817 demonstrates the validity of the query. When using a VOI congruent to CEA, the correlation increases to 0.92.

This is repeated with the spatial gene expression correlation network (*Data Set 3)* of Prkcd, Sst, Crh, Dyn and Penk, a connectivity matrix representing the voxel-wise correlation of the gene set used for this case study. BNST has again one of the strongest connections (*Figure 11 D*). *Figure 11 E, F, G* and *H* visualize the overlap of both connectivities from different perspectives demonstrating a dominant structural and genetic linkage of CEA and BNST.

"This is a post-peer-review, pre-copyedit version of an article published in NeuroInformatics. The final authenticated version is available online at: http://dx.doi.org/10.1007/s12021-019-09428-9

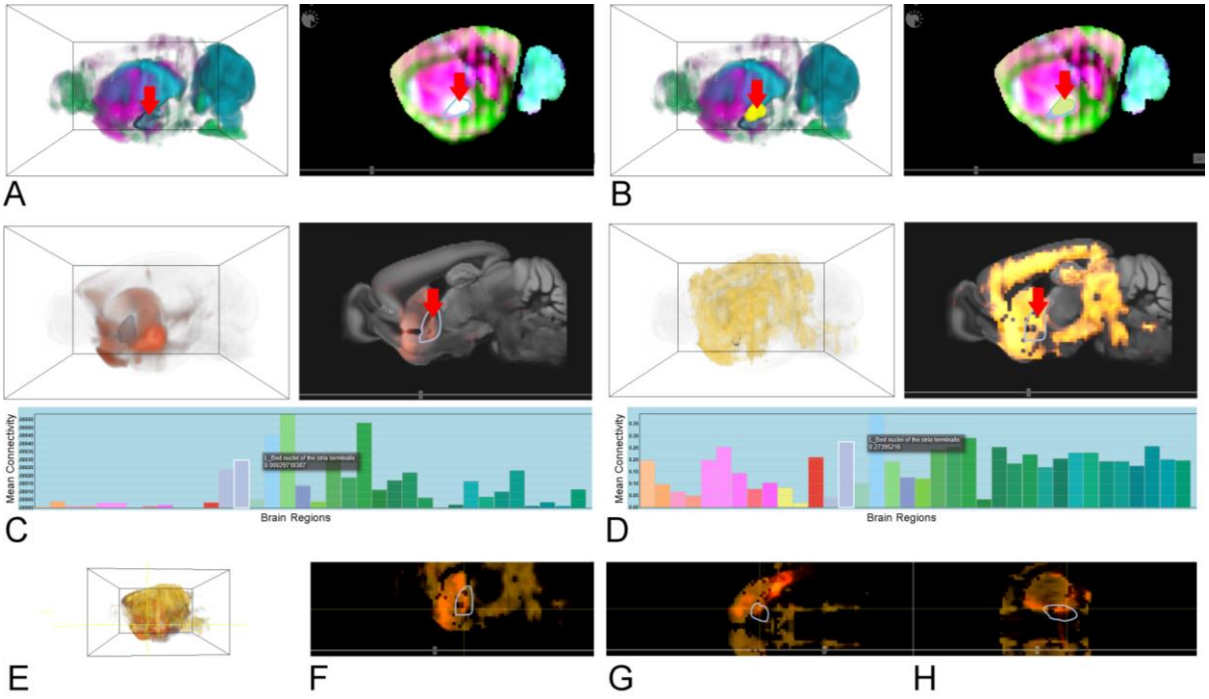


Figure 11 A: Overlap of Prkdc (cyan), Crh (green) and Dyn (purple) by aggregating the image intensity (i.e. strong overlap is white in the 2D slice view). Outlines of CEA in blue (red arrow). ***B****:* Selecting a VOI on the image overlap (yellow), ***C:*** Structural connectivity of the VOI. Outlines of BNST in dark blue (red arrow) and its connectivity profile (bars reflect mean connectivity to (Allen Brain Atlas) brain regions, with corresponding colors). ***D***: Gene-coexpression correlation of the VOI**,** analogue to *C.* ***E***: Overlap of *C* and *D* in 3D. ***F***: Overlap of *C* and *D* in a 2D slice (XY). ***G***: Overlap of *C* and *D* in a 2D slice (XZ). ***H***: Overlap of *C* and *D* in a 2D slice (YZ).

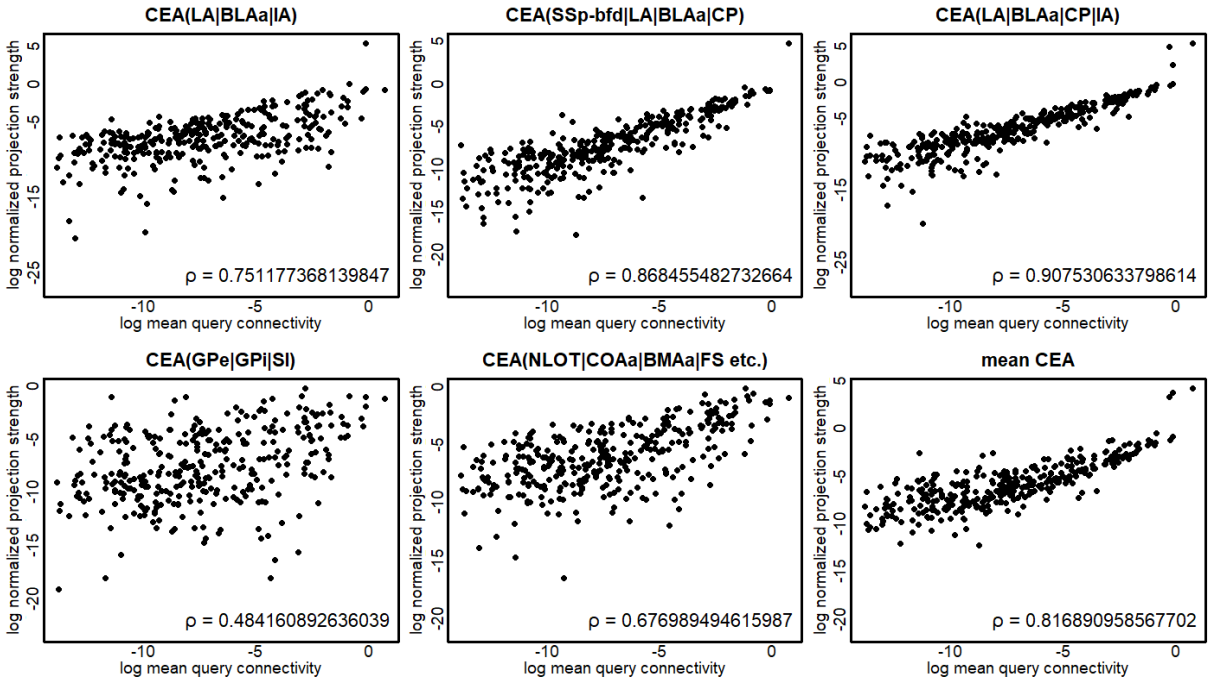


Figure 12: Correlation between mean outgoing connectivity of the query's VOI (Figure 11 B) and the normalized projection strength of sites within CEA according to Oh et. al Figure 3/Supplementary Table 2 (Oh et al. 2014).

“This is a post-peer-review, pre-copyedit version of an article published in NeuroInformatics. The final authenticated version is available online at: http://dx.doi.org/10.1007/s12021-019-09428-9

Case Study 2: Comparing networks of different modalities and species

Comparative visualization of human and animal models might be of particular interest for biomedical research and translational psychiatry. To investigate comparative functional networks across species the experts next assess this workflow by exploring functional connectivity and gene co-expression correlation from gene sets related to psychiatric traits, here exemplary autism in human (D. Li, Karnath, and Xu 2017).

For this case study we used voxel-level undirected spatial gene expression correlation (*Data Set 3*), and region-level functional connectivity (*Data Set 2*). The data is retrieved from the *Region-Wise Connectivity Database*, which highlights the usability of our data structure on different levels of hierarchical brain parcellations. An example how the user navigates these hierarchies can be seen in *Supplementary Video 2*.

To explore and compare global gene expression correlation networks and functional MRI networks across species, it is necessary to find corresponding anatomical brain regions. To our better knowledge, no comprehensive mapping of brain regions between mouse and human exists. Nevertheless, finding similarities in networks can be identified by comparing them iteratively on different anatomical levels in parallel between the two species (*Figure 13 A* and *B*, left side, colors are picked from the *AMBA* and *AHBA* and do not correspond to each other). When a user navigates the hierarchical brain parcellations (for mouse and human separately), the data structure returns the connectivity of the selected brain regions in real-time after each interaction (*Supplementary Video 2)*. Consequently, domain experts can iteratively adapt their choice of brain regions not only by their knowledge of individual inter-species region correspondence, but also based on the network similarity.

We found high coupling mostly in cortex (agranular insular and temporal association areas) and primary sensory areas (olfactory, gustatory and somatosensory areas) to the amygdala (central and medial).

Closer inspection of a subnetwork related to social behavior in autism, consisting of higher association cortex, namely insula cortex (IC), frontal pole (FP), hypothalamus (HY) and midbrain (MB), as well as the CEA, revealed gene co-expression correlation within the autism gene set was strongest within cortical regions (FP,IC) and weaker between cortex and subcortical structures (CEA,HY,MB). *Figure 13 A* shows the overlap (product) of functional connectivity and gene expression correlation of this subnetwork for mice, *Figure 13 B* is analogue for humans.

“This is a post-peer-review, pre-copyedit version of an article published in NeuroInformatics. The final authenticated version is available online at: http://dx.doi.org/10.1007/s12021-019-09428-9

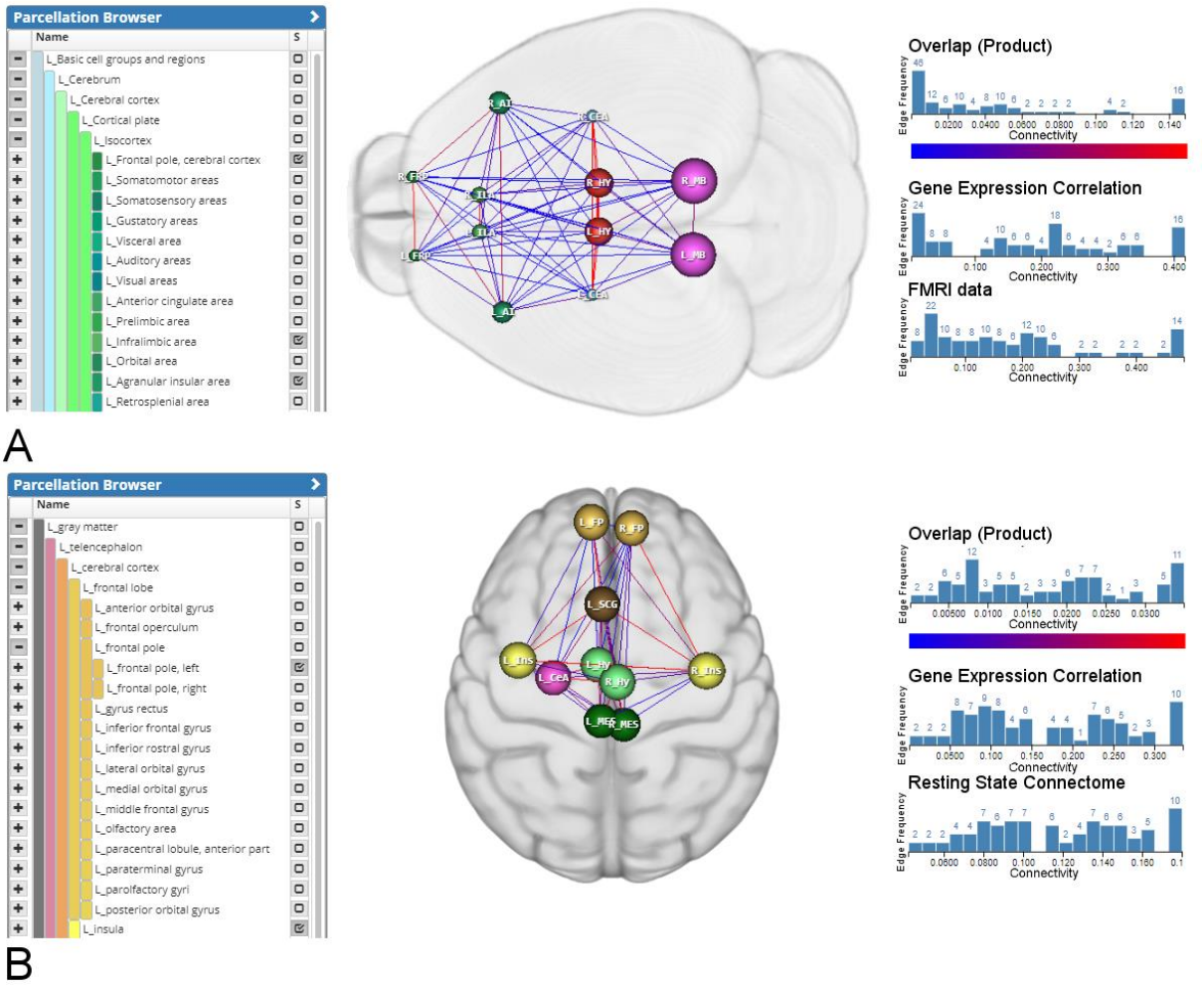


Figure 13 A: 2D mouse brain regions (green: cortical regions, red: HY, pink: MB and blue: CEA) and with overlap (product) of functional connectivity and gene expression correlation (blue: weak, red: strong) ***B***: 2D human brain regions (orange, brown and yellow: cortical regions, green: HY, dark-green: MB and pink: CEA), also with overlap (product of functional connectivity and gene expression correlation.

Discussion

We have shown that our spatial connectivity data structure outperforms state-of-the-art graph engines when querying connectivity of local brain areas. To achieve additional real-time access of outgoing/incoming connections without holding the whole connectivity matrices in the memory, a combination of data compression, spatial locality, memory mapping and hierarchical anatomical annotations is used.

Aggregation of outgoing/incoming connections of a brain area requires the reading of all edges of the involved network nodes. Therefore, row-wise data compression is used based on the specificity of the task: It reduces the total amount of data that needs to be read from the hard drive for whole rows, while it is neglectable that it is not optimized for reading single connections. As shown in the *Evaluation* section, spatial organization and sparsity of the data increases the compression factor by 2, compared to random ordering. Sparsity is often given in neurobiological connectomic data (Sporns 2016), and can be further improved by extended preprocessing of the data (Xu et al. 2015) which we suggest for future projects. One has to note that row-wise compression improves only the reading speed of rows, and therefore outgoing connections. This is not an issue for undirected connectivity graphs, since those are symmetrical (outgoing connections are equal to incoming connections), but requires a separate transposed *Connectivity Storage* for retrieving incoming connections. While this does not influence query speed, twice the disk space is needed.

A Hilbert curve is used to generate spatial locality of rows, i.e. rows, whose nodes are spatially close in the brain space are also close in the *Connectivity Storage* file. In combination with memory mapping, read-ahead paging,

"This is a post-peer-review, pre-copyedit version of an article published in NeuroInformatics. The final authenticated version is available online at: http://dx.doi.org/10.1007/s12021-019-09428-9

this greatly increases read speed, as shown in section *Evaluation*. Further, it is not necessary to hold a *Connectivity Storage* file in the memory. Therefore, one can access large matrices, with billions of edges, and execute *Aggregation Queries* on multiple matrices sequentially without loading them into memory.

New data can be imported with a REST API which creates *Connectivity Storages* and *Region-Wise Connectivity Databases* in less than an hour for connectivity matrices <100GB, scaling linearly with file size. This enables users to integrate their own, as well as data form large-scale brain initiatives on a consumer-level machine efficiently.

Depending on data acquisition techniques, neurobiological data is available on diverse scales (Betzel and Bassett 2017). To operate on different region wise levels, we used hierarchical anatomical annotations (Lein et al. 2007) and aggregated connectivity from bottom (voxel) to top (large brain regions). Since those annotations consists of only 1288 brain regions, the additional stored connections are neglectable. To bridge the gap between region and voxel levels, we created a row and column indices. These allow retrieving voxel-wise data for brain regions and mapping lower resolution data to a common reference space enabling the comparison of connectivity of different resolution. One has to note that this only represents an upsampling of the data. Since this is done in run-time, a continuous experience in visual analytics workflows is possible, for data does not need to be preprocessed. Furthermore, this technique can be used to create region-wise caches (voxel-wise outgoing/incoming connectivity of brain regions), or pyramids representations with lower resolution (voxel-wise outgoing/incoming connectivity of lower-resolution super voxels). Although these create additional storage overhead, we show in *Evaluation* that scalability is greatly improved by doing so, hence, future projects could work with even larger matrices in tera- or petabyte range.

Conclusion

In this paper, we present a novel data structure to explore heterogeneous neurobiological connectivity data of different types, modalities and scale for interactive visual analytics workflows. It enables domain experts to combine data from large-scale brain initiatives with user-generated data, by utilizing the hierarchical and spatial organization of the data. Connectivity data at different resolutions, such as mesoscale structural connectivity and region-wise functional connectivity can be queried on different levels on a common hierarchical reference space. On the lowest level, voxel-wise brain networks with billions of edges can be accessed/queried in real-time without having them loaded into working memory. It outperforms state-of-the-art graph engines in receiving connectivity of local brain areas, which allows continuous interactive exploration workflows on consumer level machines and/or via web. We demonstrate this with the implementation of a web-component for visual queries, based on *VOI* selections in 2D slice views. Results are visualized in a 3D volume rendering together with brain anatomy. Case studies conducted with domain experts showed that we could reproduce findings of neural circuits research which are currently extensively investigated experimentally. An inter-species comparison of multimodal brain networks linked to autism showed even more versatile applications, and potential use in studying psychiatric conditions.

For the future, we are aiming to extend this prototype to create a holistic framework for interactive exploration of neurobiological data. This should not only allow to access the data, but also include importing, preprocessing as well as computing network statistics in the web.

Information Sharing Statement

The implementation of the data structure is available upon individual request to the corresponding author, Florian Ganglberger or Katja Bühler.

Acknowledgments

We want to thank Florian Schulze, Nicolas Swoboda, Markus Töpfer and Emre Tosun for creating and working on parts of the web-component. This work is the result of a joint VRVis/IMP project supported by Grant 852936 of the Austrian FFG Funding Agency. VRVis is funded by BMVIT, BMWFW, Styria, SFG and Vienna Business Agency in the scope of COMET - Competence Centers for Excellent Technologies (854174) which is managed by FFG. Wulf Haubensak was supported by a grant from the European Community's Seventh Framework

“This is a post-peer-review, pre-copyedit version of an article published in NeuroInformatics. The final authenticated version is available online at: http://dx.doi.org/10.1007/s12021-019-09428-9

Programme (FP/2007-2013) / ERC grant agreement no. 311701, the Research Institute of Molecular Pathology (IMP), Boehringer Ingelheim and the Austrian Research Promotion Agency (FFG).

References

Ai, Zhiyuan, Mingxing Zhang, Yongwei Wu, Xuehai Qian, Kang Chen, and Weimin Zheng. 2017. “Squeezing out All the Value of Loaded Data: An Out-of-Core Graph Processing System with Reduced Disk I/O.” In *2017 USENIX Annual Technical Conference (USENIX ATC 17)*, 125–37. Santa Clara, CA: {USENIX} Association. https://www.usenix.org/conference/atc17/technical-sessions/presentation/ai.

Barrett, Richard, Michael Berry, Tony F Chan, James Demmel, June Donato, Jack Dongarra, Victor Eijkhout, Roldan Pozo, Charles Romine, and Henk der Vorst. 1994. *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*. SIAM.

Barthelemy, Marc. 2010. “Spatial Networks.” *Physics Reports*.

Bassett, Danielle S., and Olaf Sporns. 2017. “Network Neuroscience.” *Nature Neuroscience*. https://doi.org/10.1038/nn.4502.

Betzel, Richard F., and Danielle S. Bassett. 2017. “Multi-Scale Brain Networks.” *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2016.11.006.

Beyer, Johanna, Ali Al-Awami, Narayanan Kasthuri, Jeff W. Lichtman, Hanspeter Pfister, and Markus Hadwiger. 2013. “ConnectomeExplorer: Query-Guided Visual Analysis of Large Volumetric Neuroscience Data.” *IEEE Transactions on Visualization and Computer Graphics*. https://doi.org/10.1109/TVCG.2013.142.

Chen, Yiqi, Zhiyuan Lin, Robert Pienta, Minsuk Kahng, and Duen Horng Chau. 2015. “Towards Scalable Graph Computation on Mobile Devices.” In *Proceedings - 2014 IEEE International Conference on Big Data, IEEE Big Data 2014*. https://doi.org/10.1109/BigData.2014.7004353.

Chi, Yuze, Guohao Dai, Yu Wang, Guangyu Sun, Guoliang Li, and Huazhong Yang. 2016. “NXgraph: An Efficient Graph Processing System on a Single Machine.” In *2016 IEEE 32nd International Conference on Data Engineering, ICDE 2016*. https://doi.org/10.1109/ICDE.2016.7498258.

Demir, Engin, and Cevdet Aykanat. 2010. “Efficient Successor Retrieval Operations for Aggregate Query Processing on Clustered Road Networks.” *Information Sciences*. https://doi.org/10.1016/j.ins.2010.03.015.

Essen, David C Van, Stephen M Smith, Deanna M Barch, Timothy E J Behrens, Essa Yacoub, Kamil Ugurbil, WU-Minn H C P Consortium, and others. 2013. “The WU-Minn Human Connectome Project: An Overview.” *Neuroimage* 80. Elsevier: 62–79.

Ganglberger, Florian, Joanna Kaczanowska, Josef M. Penninger, Andreas Hess, Katja Bühler, and Wulf Haubensak. 2017. “Predicting Functional Neuroanatomical Maps from Fusing Brain Networks with Genetic Information.” *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2017.08.070.

Garulli, Luca. 2010. “OrientDB.” Orient Technologies LTD. 2010.

Han, Wook-Shin, Sangyeon Lee, Kyungyeol Park, Jeong-Hoon Lee, Min-Soo Kim, Jinha Kim, and Hwanjo Yu. 2013. “TurboGraph: A Fast Parallel Graph Engine Handling Billion-Scale Graphs in a Single PC.” *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. https://doi.org/10.1145/2487575.2487581.

Haubensak, Wulf, Prabhat S Kunwar, Haijiang Cai, Stephane Ciocchi, Nicholas R Wall, Ravikumar Ponnusamy, Jonathan Biag, et al. 2010. “Genetic Dissection of an Amygdala Microcircuit That Gates Conditioned Fear.” *Nature* 468 (7321). Nature Publishing Group: 270–76.

Hawrylycz, Michael J, Ed S Lein, Angela L Guillozet-Bongaarts, Elaine H Shen, Lydia Ng, Jeremy A Miller, Louie N van de Lagemaat, et al. 2012. “An Anatomically Comprehensive Atlas of the Adult Human Brain Transcriptome.” *Nature* 489 (7416). Nature Publishing Group: 391–99. http://dx.doi.org/10.1038/nature11405.

Hilbert, David. 1891. “Ueber Die Stetige Abbildung Einer Line Auf Ein Flächenstück.” *Mathematische Annalen*

“This is a post-peer-review, pre-copyedit version of an article published in NeuroInformatics. The final authenticated version is available online at: http://dx.doi.org/10.1007/s12021-019-09428-9

38 (3). Springer: 459–60.

Kim, Joshua, Xiangyu Zhang, Shruti Muralidhar, Sarah A LeBlanc, and Susumu Tonegawa. 2017. “Basolateral to Central Amygdala Neural Circuits for Appetitive Behaviors.” *Neuron* 93 (6). Elsevier: 1464–79.

Kyrola, Aapo, Guy Blelloch, and Carlos Guestrin. 2012. “GraphChi: Large-Scale Graph Computation on Just a PC Disk-Based Graph Computation.” *Proceedings of the 10th USENIX Conference on Operating Systems Design and Implementation*. https://doi.org/10.1109/HPCA.2015.7056066.

Lein, Ed S, Michael J Hawrylycz, Nancy Ao, Mikael Ayres, Amy Bensinger, Amy Bernard, Andrew F Boe, et al. 2007. “Genome-Wide Atlas of Gene Expression in the Adult Mouse Brain.” *Nature* 445 (7124): 168–76. https://doi.org/10.1038/nature05453.

Leskovec, Jure, and Rok Sosič. 2016. “SNAP: A General-Purpose Network Analysis and Graph-Mining Library.” *ACM Transactions on Intelligent Systems and Technology*. https://doi.org/10.1145/2898361.

Li, Dongyun, Hans Otto Karnath, and Xiu Xu. 2017. “Candidate Biomarkers in Children with Autism Spectrum Disorder: A Review of MRI Studies.” *Neuroscience Bulletin*. https://doi.org/10.1007/s12264-017-0118-1.

Li, Kaiming, Lei Guo, Carlos Faraco, Dajiang Zhu, Hanbo Chen, Yixuan Yuan, Jinglei Lv, et al. 2012. “Visual Analytics of Brain Networks.” *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2012.02.075.

Lin, Zhiyuan, Duen Horng Polo Chau, and U. Kang. 2013. “Leveraging Memory Mapping for Fast and Scalable Graph Computation on a PC.” In *Proceedings - 2013 IEEE International Conference on Big Data, Big Data 2013*. https://doi.org/10.1109/BigData.2013.6691739.

Lin, Zhiyuan, Minsuk Kahng, Kaeser Md Sabrin, Duen Horng Polo Chau, Ho Lee, and U. Kang. 2014. “MMap: Fast Billion-Scale Graph Computation on a PC via Memory Mapping.” In *Proceedings - 2014 IEEE International Conference on Big Data, IEEE Big Data 2014*. https://doi.org/10.1109/BigData.2014.7004226.

Macko, Peter, Virendra J. Marathe, Daniel W. Margo, and Margo I. Seltzer. 2015. “LLAMA: Efficient Graph Analytics Using Large Multiversioned Arrays.” In *Proceedings - International Conference on Data Engineering*. https://doi.org/10.1109/ICDE.2015.7113298.

Markram, Henry, Karlheinz Meier, Thomas Lippert, Sten Grillner, Richard Frackowiak, Stanislas Dehaene, Alois Knoll, et al. 2011. “Introducing the Human Brain Project.” In *Procedia Computer Science*. https://doi.org/10.1016/j.procs.2011.12.015.

Oh, Seung Wook, Julie A. Harris, Lydia Ng, Brent Winslow, Nicholas Cain, Stefan Mihalas, Quanxin Wang, et al. 2014. “A Mesoscale Connectome of the Mouse Brain.” *Nature* 508 (7495): 207–14. http://www.nature.com/doifinder/10.1038/nature13186.

Papadias, Dimitris, J Zhang, Nikos Mamoulis, and Y Tao. 2003. “Query Processing in Spatial Network Databases.” *Proceedings of the 29th International Conference on Very Large Data Bases*. https://doi.org/10.1016/B978-012722442-8/50076-8.

Pienta, Robert, James Abello, Minsuk Kahng, and Duen Horng Chau. 2015. “Scalable Graph Exploration and Visualization: Sensemaking Challenges and Opportunities.” In *2015 International Conference on Big Data and Smart Computing, BIGCOMP 2015*. https://doi.org/10.1109/35021BIGCOMP.2015.7072812.

Poo, Mu ming, Jiu lin Du, Nancy Y. Ip, Zhi Qi Xiong, Bo Xu, and Tieniu Tan. 2016. “China Brain Project: Basic Neuroscience, Brain Diseases, and Brain-Inspired Computing.” *Neuron*. https://doi.org/10.1016/j.neuron.2016.10.050.

Radke, Anna K. 2009. “The Role of the Bed Nucleus of the Stria Terminalis in Learning to Fear.” *Journal of Neuroscience* 29 (49). Soc Neuroscience: 15351–52.

Richiardi, Jonas, and Andre Altmann. 2015. “Correlated Gene Expression Supports Synchronous Activity in Brain Networks.” *Science* 348 (6240): 11–14.

Roy, Amitabha, Ivo Mihailovic, and Willy Zwaenepoel. 2013. “X-Stream: Edge-Centric Graph Processing Using Streaming Partitions.” *Proceedings of the ACM Symposium on Operating Systems Principles*. https://doi.org/10.1145/2517349.2522740.

Saalfeld, Stephan, Albert Cardona, Volker Hartenstein, and Pavel Tomančák. 2009. “CATMAID: Collaborative

“This is a post-peer-review, pre-copyedit version of an article published in NeuroInformatics. The final authenticated version is available online at: http://dx.doi.org/10.1007/s12021-019-09428-9

Annotation Toolkit for Massive Amounts of Image Data.” *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btp266.

Shekhar, Shashi, and Duen Ren Liu. 1997. “CCAM: A Connectivity-Clustered Access Method for Networks and Network Computations.” *IEEE Transactions on Knowledge and Data Engineering*. https://doi.org/10.1109/69.567054.

Sherbondy, Anthony, David Akers, Rachel Mackenzie, Robert Dougherty, and Brian Wandell. 2005. “Exploring Connectivity of the Brain’s White Matter with Dynamic Queries.” In *IEEE Transactions on Visualization and Computer Graphics*. https://doi.org/10.1109/TVCG.2005.59.

Sporns, Olaf. 2016. “Connectome Networks: From Cells to Systems.” In *Research and Perspectives in Neurosciences*. https://doi.org/10.1007/978-3-319-27777-6_8.

Tauheed, Farhan, Sadegh Nobari, Laurynas Biveinis, Thomas Heinis, and Anastasia Ailamaki. 2013. “Computational Neuroscience Breakthroughs through Innovative Data Management.” In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-642-40683-6_2.

Xu, Ting, Zhi Yang, Lili Jiang, Xiu Xia Xing, and Xi Nian Zuo. 2015. “A Connectome Computation System for Discovery Science of Brain.” *Science Bulletin*. https://doi.org/10.1007/s11434-014-0698-3.

Young, L J, and Z Wang. 2004. “The Neurobiology of Pair Bonding.” *Nat Neurosci* 7 (10): 1048–54. http://www.ncbi.nlm.nih.gov/pubmed/15452576.

Zheng, Da, Disa Mhembere, Randal Burns, Joshua Vogelstein, Carey E Priebe, and Alexander S Szalay. 2015. “FlashGraph: Processing Billion-Node Graphs on an Array of Commodity SSDs.” In *Proceedings of the 13th USENIX Conference on File and Storage Technologies*, 45–58. FAST’15. Berkeley, CA, USA: USENIX Association. http://dl.acm.org/citation.cfm?id=2750482.2750486.

Supplementary Note 1: Creation of Connectivity Matrices

1. Structural Connectivity for Mouse

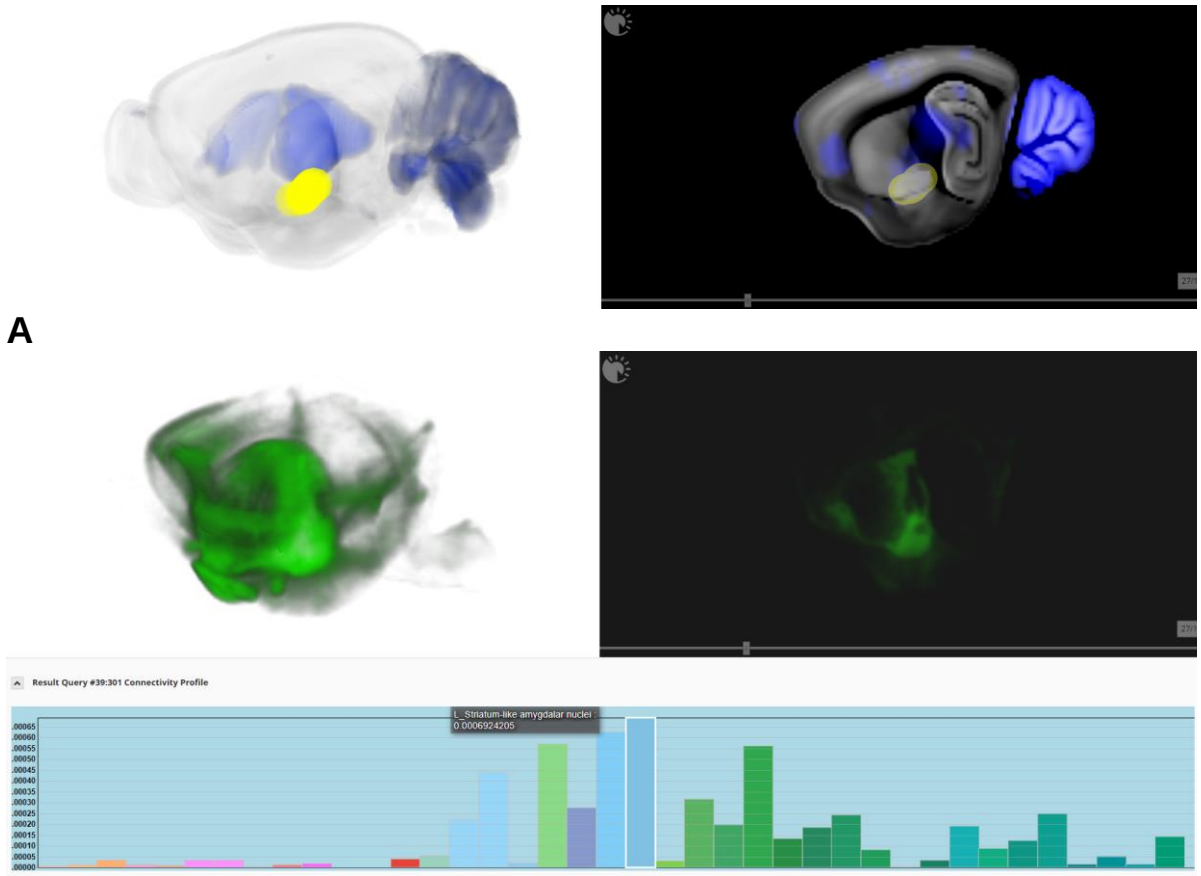
We created the voxel level mouse connectome in a similar way as in Ganglberger et. Al (Ganglberger et al., 2017). The mouse connectome was retrieved from the Allen Brain Atlas API as (structural) connectivity from all 2173 available injection sites (state May 2018) to their target sites given as image data detailing projections labeled by rAAV tracers via serial two-photon tomography (Oh et al., 2014). Injection sites are given by coordinates in a 100 micron mouse brain space (132x80x114), where the rAAV tracers had been injected, and an injection volume, depicting the volume around the injection site affected by the tracer. Therefore, the connectivity for a single injection site defines the connectivity of all voxels within its injection volume. So for every voxel in the mouse brain space (source voxel) we took the connectivity from all covering injection volumes to all other voxels in brain space (target voxels). For overlapping injection volumes, we took the maximum connectivity for each target voxel. Since the majority of injection sites are located on the right hemisphere, we compensate this by mirroring them to the left hemisphere. This "inflates" the connectivity from 2173 injections sites to a 67500 x 450000 connectome (67500 source voxels with 450000 target voxels. In order to increase the sparsity of the data, the connectome was threshold according to Oh, S. W. *et al.* (Oh et al., 2014), Extended Data Figure 7, which minimizes the amount of false positive connections. A Matlab script for downloading the connections from AMBA, an R script for building the connectome, as well as the AMBA connectome are provided on request.

2. Spatial gene expression correlation networks

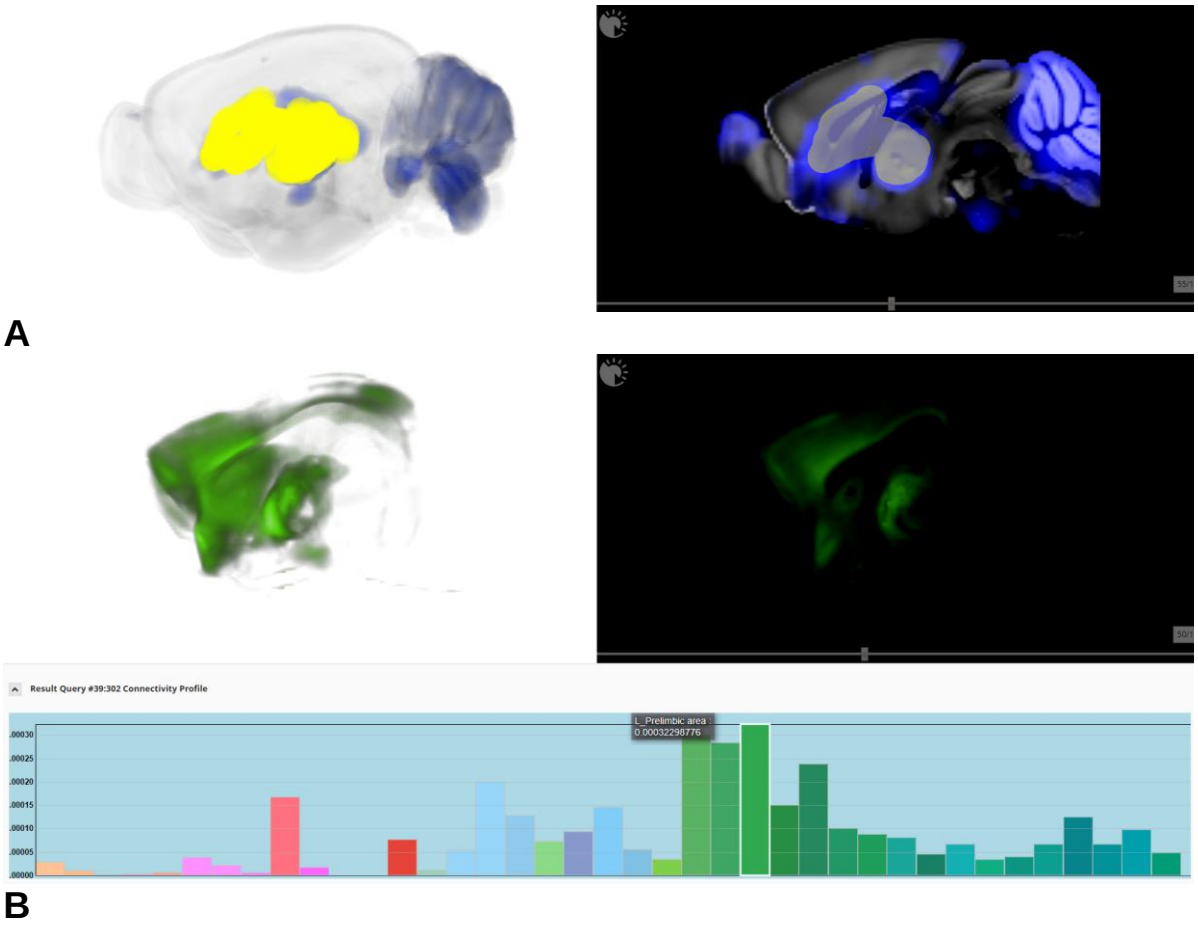
Mouse gene expression has been retrieved from the Allen Mouse Brain Atlas *(Lein et al., 2007)* on 200 micron voxel-level resolution (i.e. for every gene we retrieved 200 micron resolution image data). To create spatial gene expression correlation networks for a gene set, we first normalized the gene expression data according to Ganglberger et al (Ganglberger et al., 2017). The spatial gene expression correlation between two voxels is then computed by the pearson correlation coefficient within the gene set.

For human, we retrieved region wise gene expression from the Allen Human Brain Atlas (Hawrylycz et al., 2012). The Allen Institute provides an affine transformation to MNI152 (Fonov et al., 2011) space by its API. We used resting state functional connectivity from the Human Connectome Project (Glasser et al., 2013), which is also in MNI152space (Fonov et al., 2011). Data normalization was performed in a robust way (median/mad) since fewer data points are more sensitive to outliers compared to mouse data.

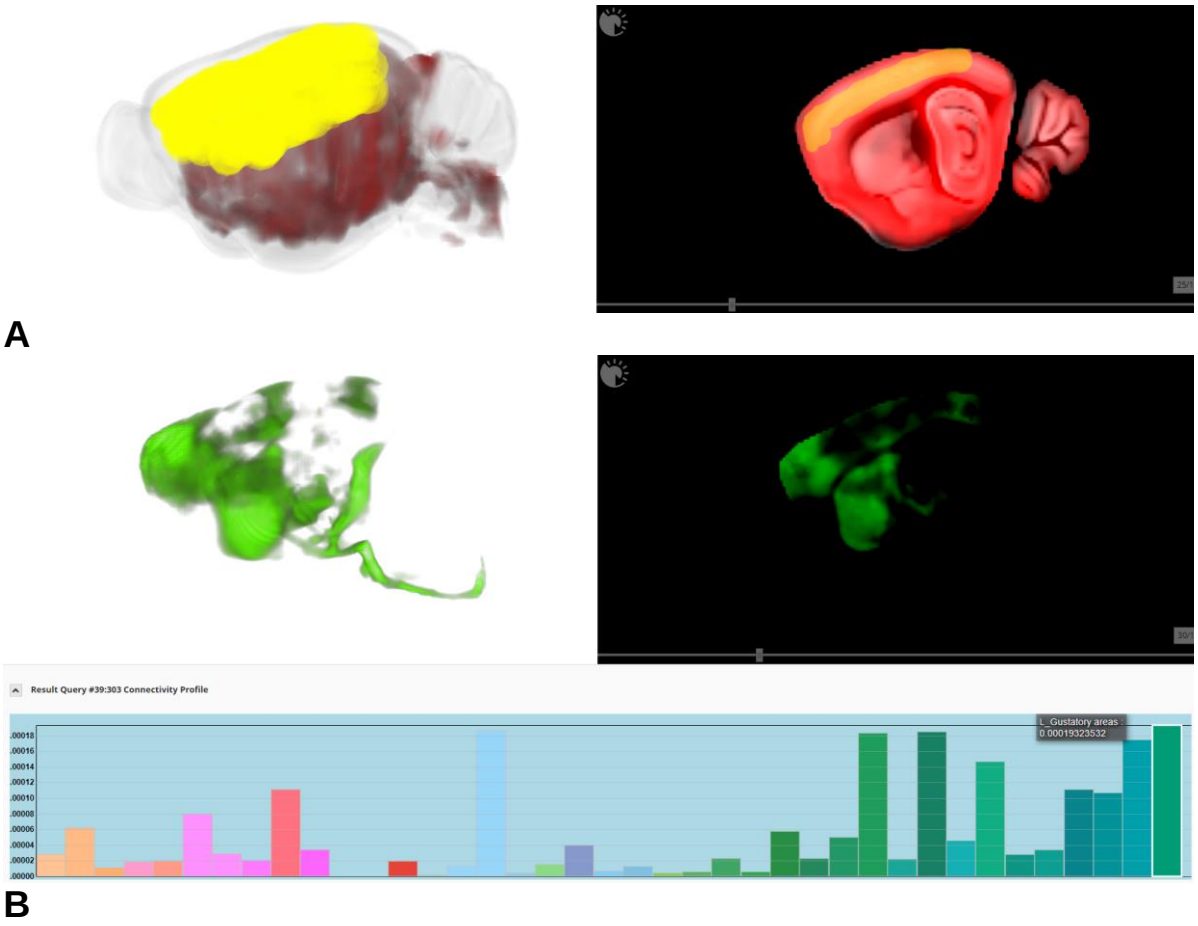
Supplementary Note 2: Manual User Queries



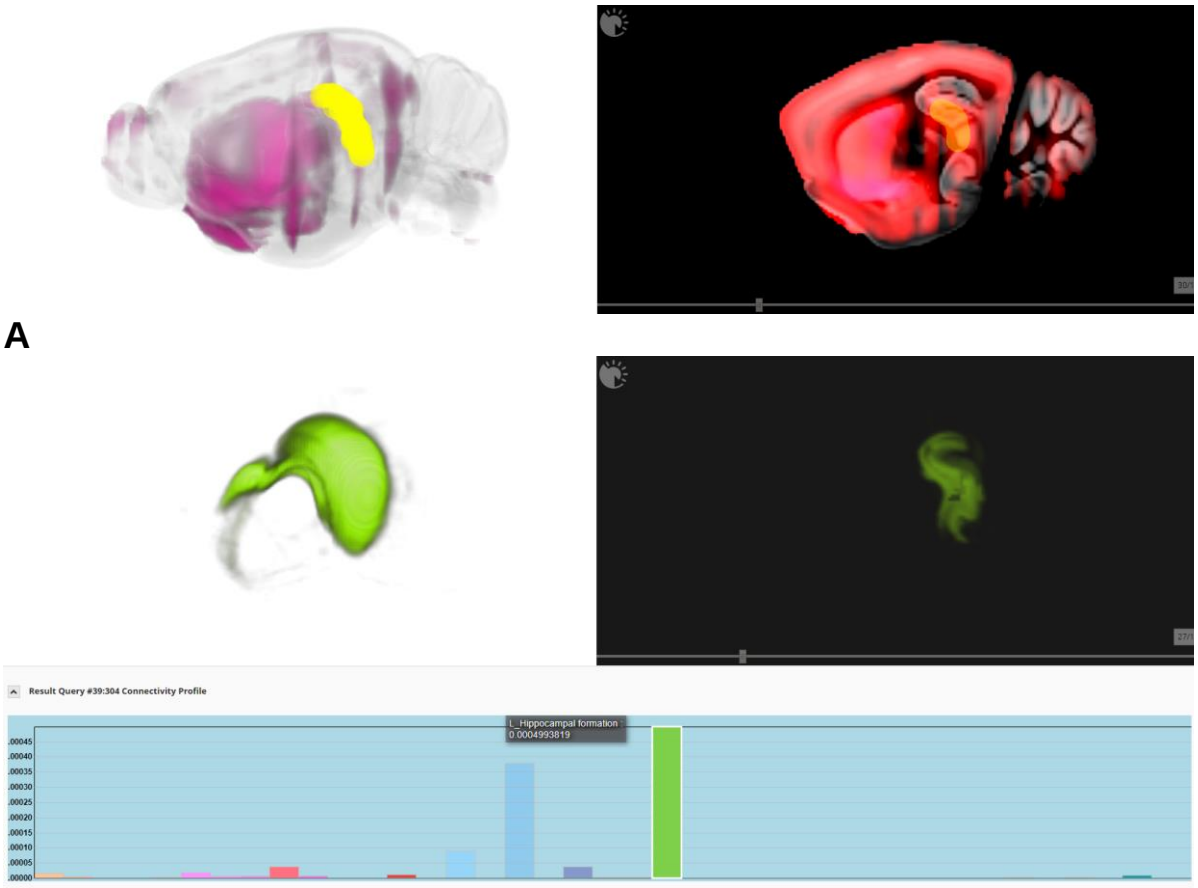
Supplementary Figure 1 A: Gene expression of PKCD with a selected *VOI* (yellow). **B:** Aggregated structural connectivity (targets) of the *VOI* as well as its connectivity profile.



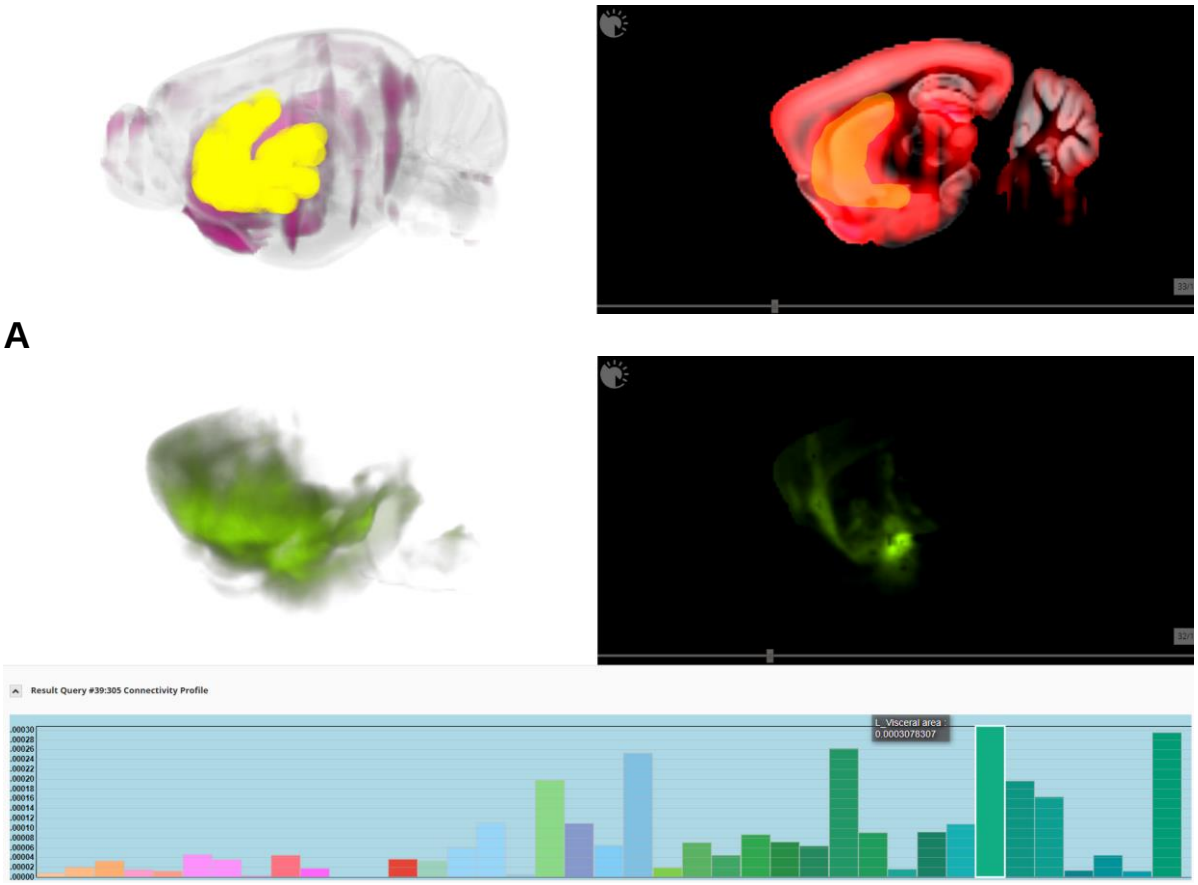
Supplementary Figure 2 A: Gene expression of PKCD with a selected *VOI* (yellow). **B:** Aggregated structural connectivity (targets) of the *VOI* as well as its connectivity profile.



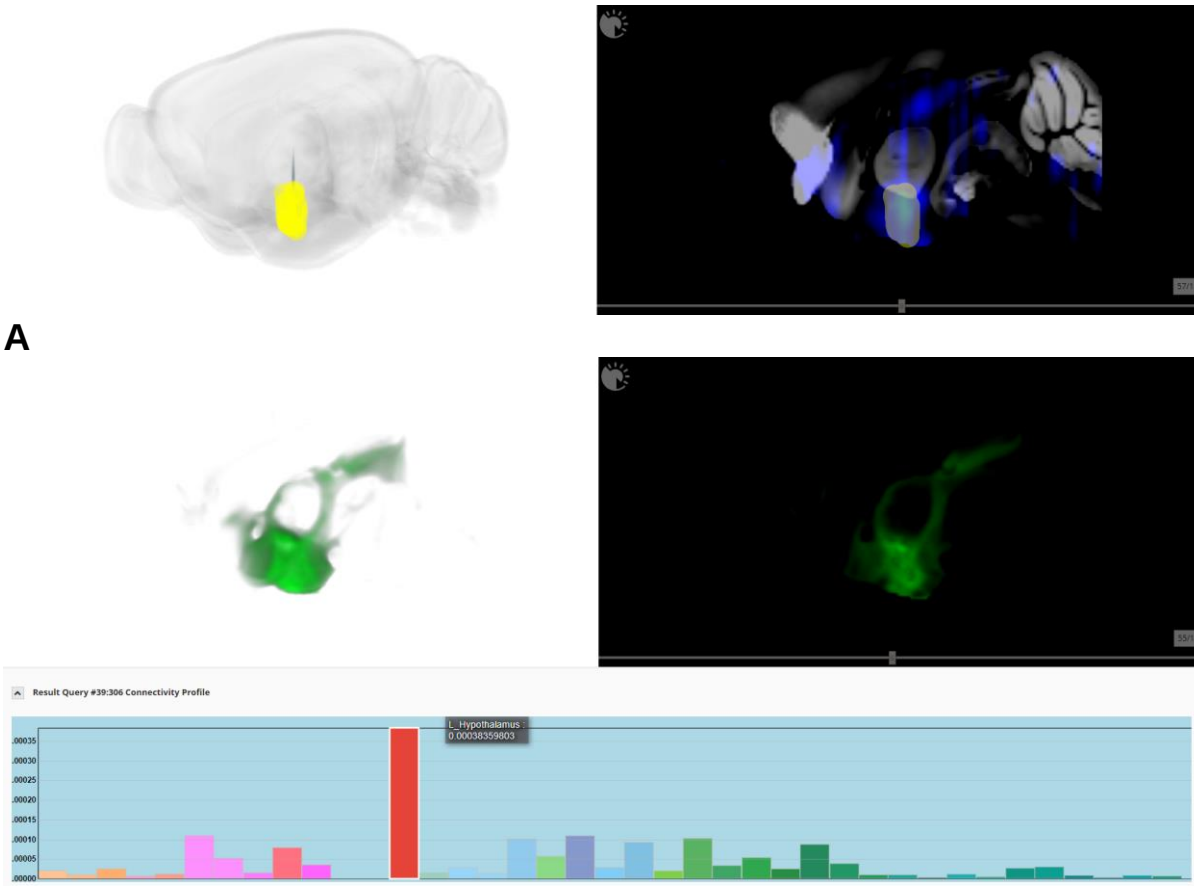
Supplementary Figure 3 A: Gene expression of SST with a selected *VOI* (yellow). **B:** Aggregated structural connectivity (targets) of the *VOI* as well as its connectivity profile.



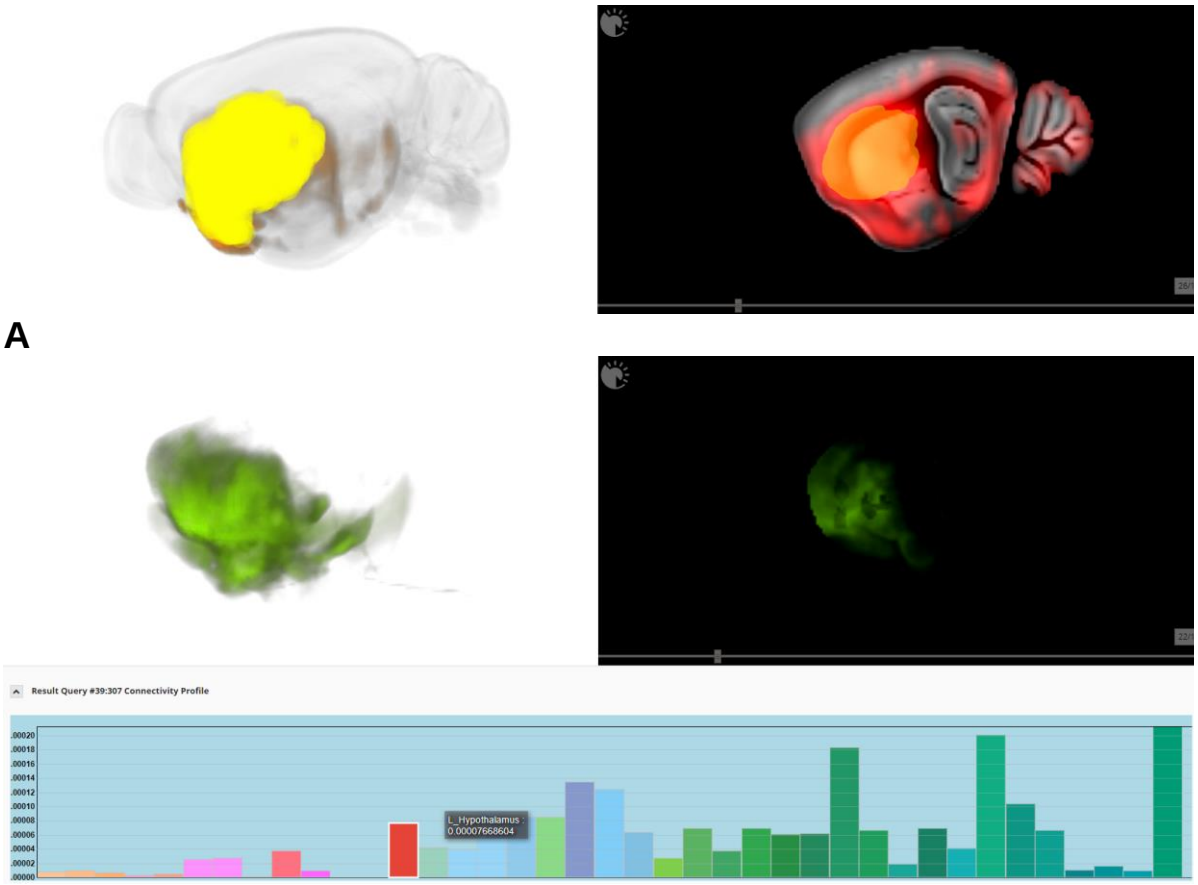
Supplementary Figure 4 A: Gene expression of PDYN with a selected *VOI* (yellow). **B:** Aggregated structural connectivity (targets) of the *VOI* as well as its connectivity profile



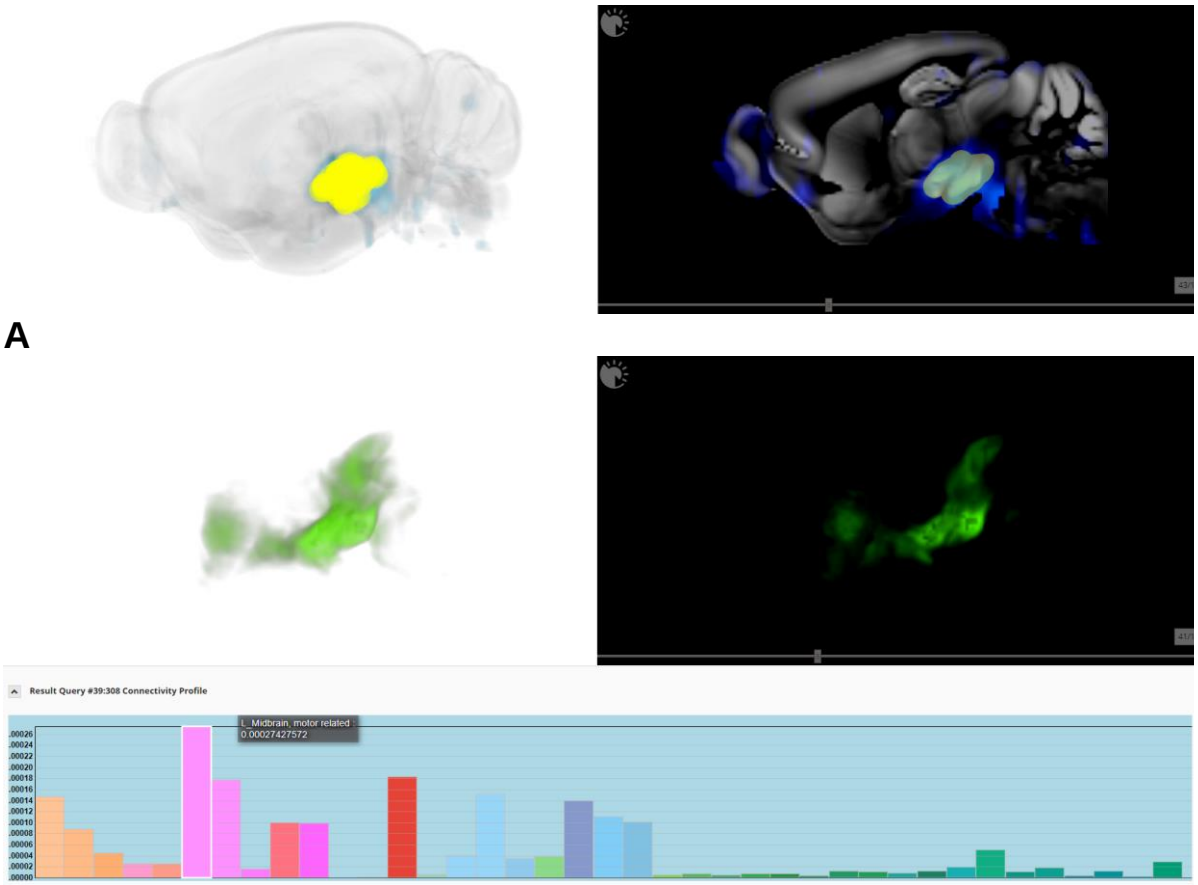
Supplementary Figure 5 A: Gene expression of PKCD with a selected *VOI* (yellow). **B:** Aggregated structural connectivity (targets) of the *VOI* as well as its connectivity profile.



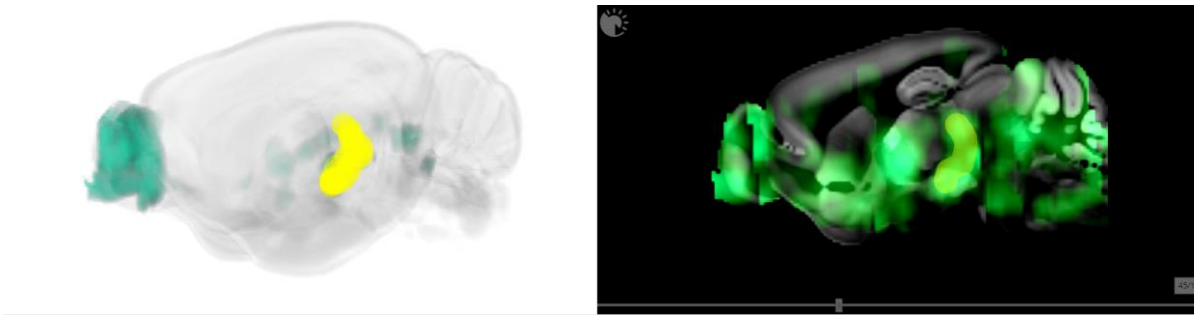
Supplementary Figure 6 A: Gene expression of Oxytocin with a selected *VOI* (yellow). **B:** Aggregated structural connectivity (targets) of the *VOI* as well as its connectivity profile.



Supplementary Figure 7 A: Gene expression of D1R with a selected *VOI* (yellow). **B:** Aggregated structural connectivity (targets) of the *VOI* as well as its connectivity profile.



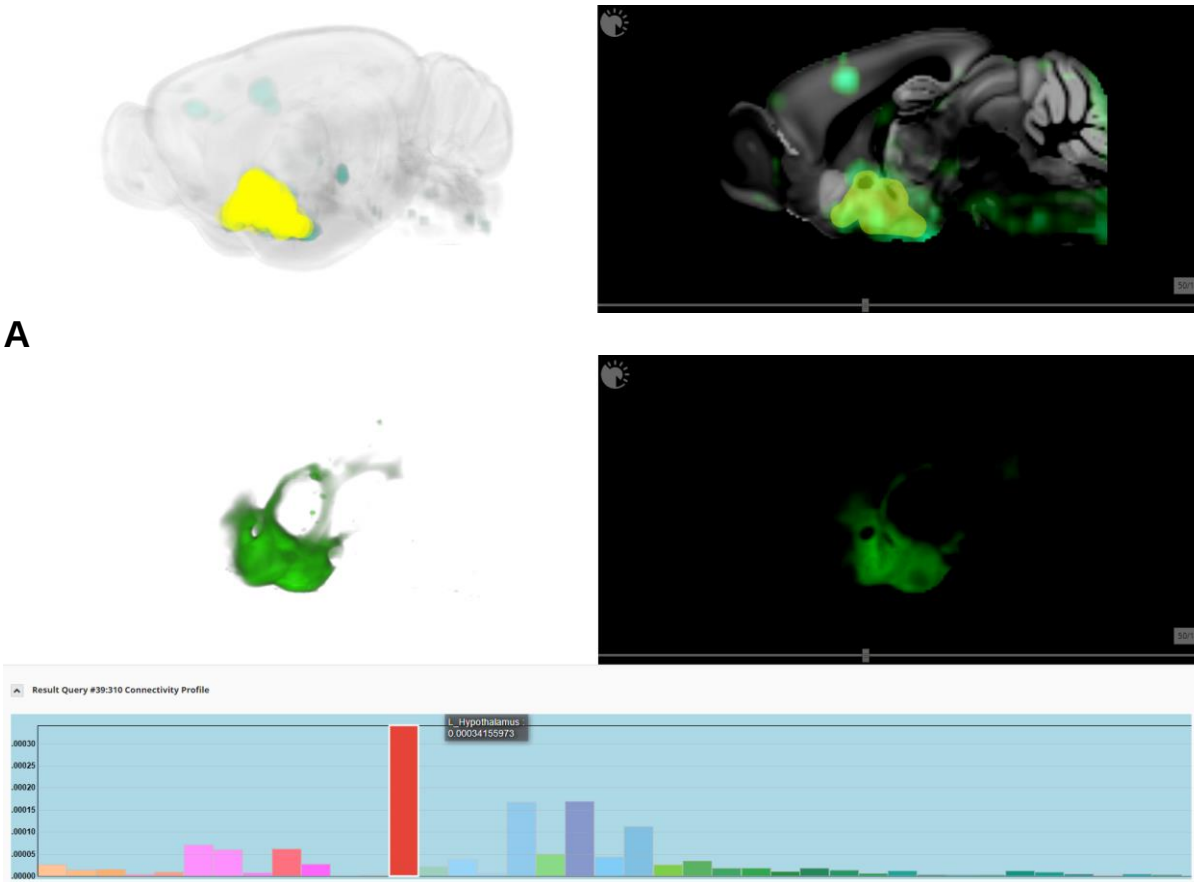
Supplementary Figure 8 A: Gene expression of Slc6a3 with a selected *VOI* (yellow). **B:** Aggregated structural connectivity (targets) of the *VOI* as well as its connectivity profile.



A

B

Supplementary Figure 9 A: Gene expression of TH with a selected *VOI* (yellow). **B:** Aggregated structural connectivity (targets) of the *VOI* as well as its connectivity profile. **C:** Aggregated structural connectivity in 3D (without *VOI*)

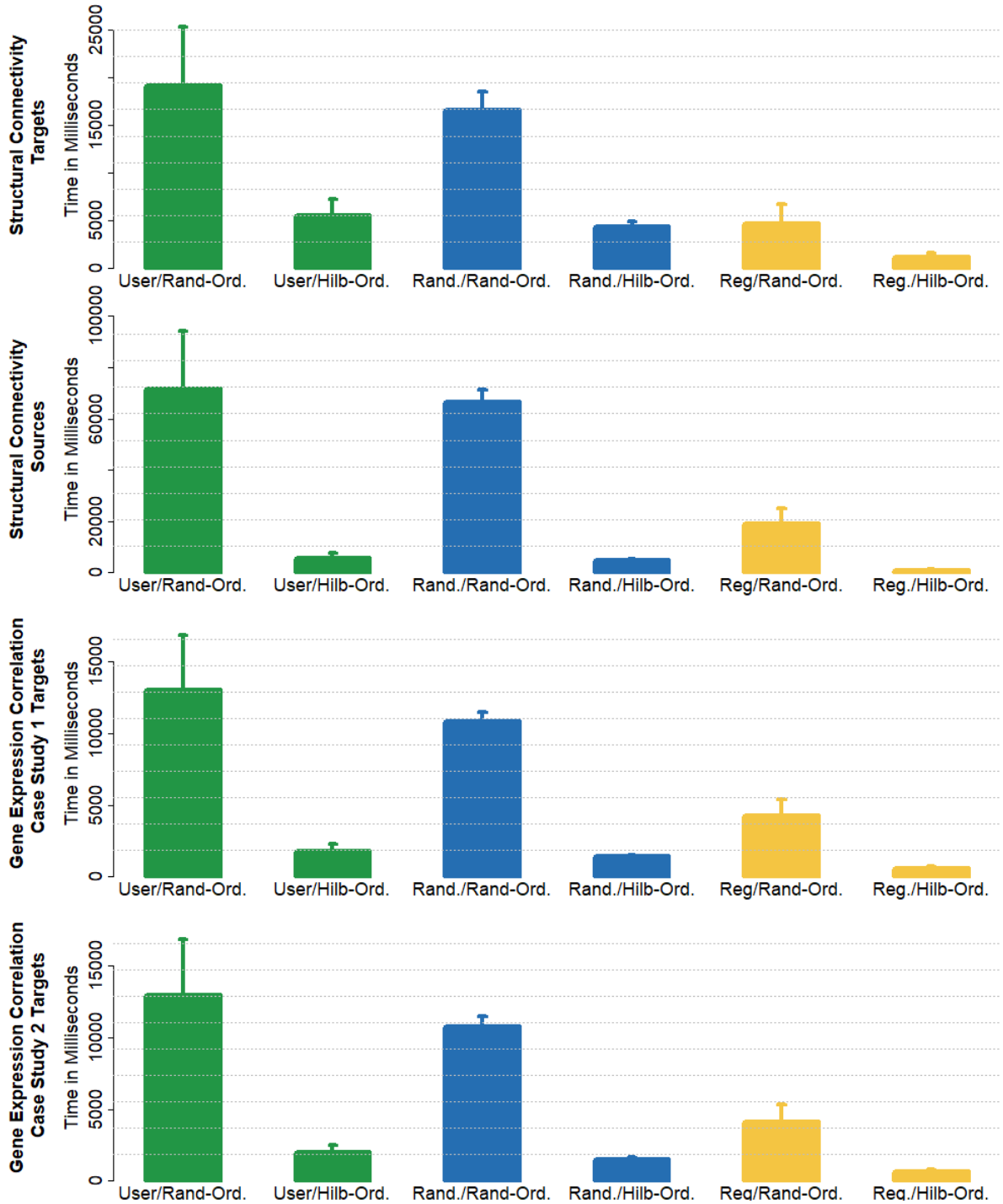


Supplementary Figure 1 0A: Gene expression of AVP with a selected *VOI* (yellow). **B:** Aggregated structural connectivity (targets) of the *VOI* as well as its connectivity profile.

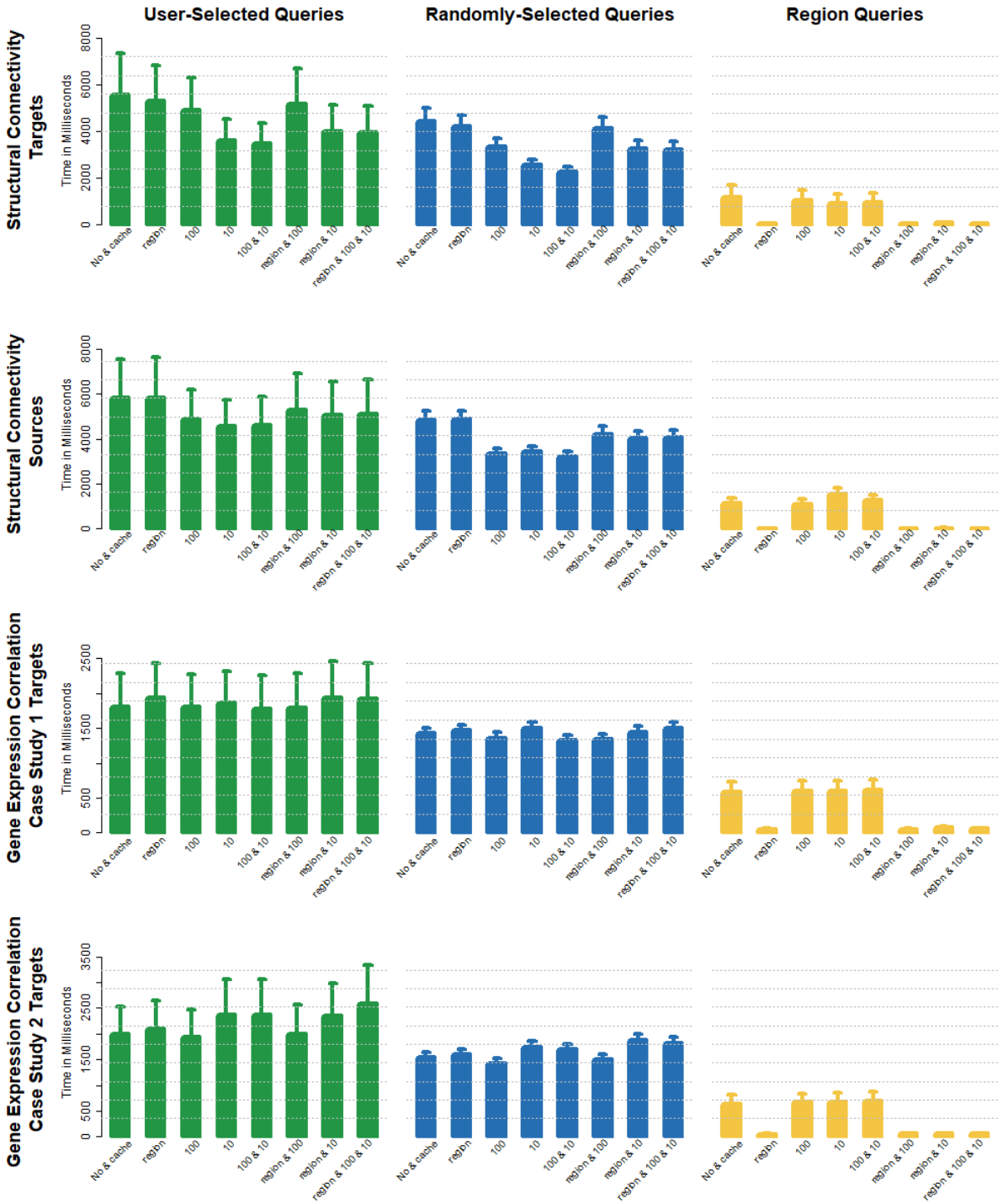
Supplementary Note 3: Brain Regions

We used following regions taken from the Allen Mouse Brain Atlas (Lein et al., 2007) (direct link to atlas: http://atlas.brain-map.org/atlas?atlas=2): Nucleus acumbens, Frontal pole, Visual areas, Ventral posteromedial nucleus, Substantia nigra, Parvicellular reticular nucleus, Vermal regions, Lateral hypothalamic area, Dentate gyrus and Accessory olfactory bulb

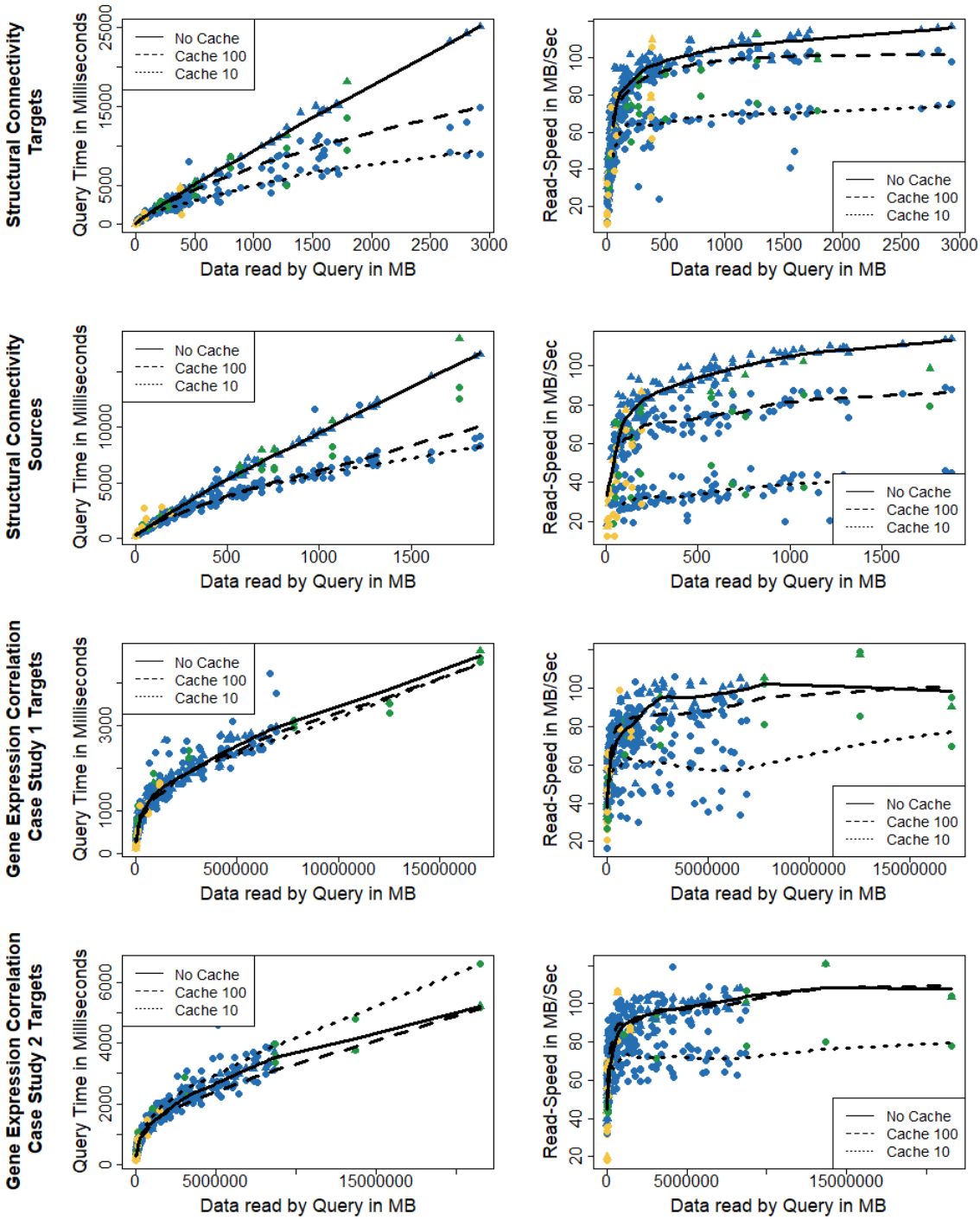
Supplementary Note 4: Evaluation on Hard Disk



Supplementary Figure 11: Effect of spatial-ordering on query-speed on different connectivity matrices. Bars show the mean query-time with standard error of 10 user-defined *VOI* queries (green), 100 random *VOI* queries (blue) and 10 region *VOI* queries (yellow), for Hilbert-Ordering and Random-Ordering.



Supplementary Figure 12: Effect of *Connectivity Cache* on query-speed on different connectivity matrices. Bars show the mean query-time with standard error of 10 user-defined VOI queries (green), 100 random *VOI* queries (blue) and 10 region *VOI* queries (yellow), for different types of caches and their combination.



Supplementary Figure 13: Relation of query-time and read-speed on query size for different *Connectivity Matrices, Connectivity Cache*s and query types. The left column shows the query time vs query size for queries executed with cache (●) and without (▲), while the color depicting the query type (green=user query, blue=random query and yellow=region query). The right column shows read speed on the *Connectivity Storage Files* vs query size, similarly encoded. LOWESS regression lines are added to see the overall trend for different cache sizes.

Supplementary Note 5: Implementation of State-of-the-Art Tools for Aggregation Queries

1. FlashGraph

To evaluate *FlashGraph* (Zheng et al., 2015), we used the *FlashGraph* C++ API which is provided in a GitHub repository (https://github.com/flashxio/FlashX). *FlashGraph* uses an edgelist format (source_node_id target_node_id float_value) which we created similarly to *Supplementary Note 1* (the matrices are equal but not in the connectivity matrix format). "Zero"-edges (edges without connectivity) were excluded. The files have a size of 94 GB (SC), 23 GB (CS1) and 27 GB (CS2). To access these matrices by *FlashGraph*, the data need to be converted to a special file format, which splits the edges into an adjacency file and an index file. In this graph format, every voxel in brain space represents a vertex (node) of a graph, while its edges show the connectivity between them. For this task, *FlashGraph* provides a command-line tool el2fg (= "edge list to flash graph"), which is also provided in the repository. We used the single-precision float-point format to encode weights as 4 bytes to create 3 adjacency/index files: SC: 54GB, CS1: 18 GB, CS2: 18GB. Nodes in the edgelist where ordered with a Hilbert curve, so FlashGraph's reading of the adjacency files profits also from read-ahead-paging for local queries.

To perform *Aggregation Queries* we created a special Aggregation Vertex Class (see FlashGraph documentation at https://flashxio.github.io/FlashX-doc/FlashGraph-user-guide.html for details on Vertex Classes), shown in Supplementary Figure 14. An *Aggregation Query* is then performed by iterating over all vertices of the *VOI* (i.e. all voxels of the *VOI*), so only their edge data needs to be loaded. Since only data that is relevant for the query is loaded, this is more efficient than accessing the entire graph. We then summarizing the outgoing connectivity of the *VOI* vertices in all vertices connected to them (i.e. a vertex stores the sum of all receiving edges from the *VOI* vertices). This does not require IO operations, since the vertex state (without edge data) is kept in memory by *FlashGraph*. This resulting connectivity can then be retrieved by iterating over all vertices without loading edge data. For simplicity, retrieved query results were in their original resolution and were not upsampled to the 100-micron standard brain (so the queries are in fact about 100 milliseconds slower). Keeping the vertex state independent of its edge data (which is only loaded on demand by the user), queries can be performed much faster compared to *GraphChi* (Kyrola, Blelloch, & Guestrin, 2012).

```
class aggregation_vertex: public compute_directed_vertex {
    public:
        float aggregatedConnectivity=0; //aggregated connectivity for every vertex
        static bool loadTargets; //if targets or sources should be loaded

        aggregation_vertex(vertex_id_t id): compute_directed_vertex(id) {
        }

        void run(vertex_program &prog) {
            //request vertices
            //class should be initialized only with VOI vertices
            vertex_id_t id = prog.get_vertex_id(*this);
            request_vertices(&id, 1);
        }

        void run(vertex_program &prog, const page_vertex &vertex) {
            //get all outgoing or incoming edges (targets or sources)
            edge_seq_iterator neigh_it = vertex.get_neigh_seq_it(loadTargets?OUT_EDGE:IN_EDGE);

            // Iterator for egde count (weight) attribute
            safs::page_byte_array::seq_const_iterator<edge_count> count_it =
```

```
                ((const
page_directed_vertex&)(vertex)).get_data_seq_it<edge_count>(loadTargets?OUT_EDGE:IN_EDGE);

            while (neigh_it.has_next()) {
                vertex_id_t nid = neigh_it.next();
                edge_count e = count_it.next();
                unsigned val = e.get_count();

                //get target/source vertex and aggregate the connectivity there
                aggregation_vertex& v = (aggregation_vertex&) prog.get_graph().get_vertex(nid);
                v.aggregatedConnectivity+=reinterpret_cast<float &>(val);

            }
        }

        void run_on_vertex(vertex_program &prog, page_vertex &vertex) {

        }

        float get_result(){
            return aggregatedConnectivity;
        }
};
```

***Supplementary Figure 14*:** Aggregation vertex class which is used by the *FlashGraph* API to execute *Aggregation Queries*

2. GraphChi

We implemented *Aggregation Queries* for GraphChi (Kyrola et al., 2012) by using the *GraphChi* C++ API, from the *GraphChi* GitHub repository (https://github.com/GraphChi). Similarly to *FlashGraph*, *GraphChi* uses the similar edgelist format (source_node_id target_node_id float_value) as input data, so we used the same files. To load the data with GraphChi, the data is automatically split by the GraphChi into sub-partitions, so called “shards” (Kyrola et al., 2012) before queries can be executed (shard space for SC: 19GB, CS1: 7GB, CS2: 8GB). The number of shards has been determined automatically by GraphChi, but was also tested with 100, 100 and 1000, which showed no performance improvements. This is consistent with the performance evaluation in the GraphChi paper (Kyrola et al., 2012), Figure 8.

We implemented *Aggregation Queries* in an Aggregation GraphChi Program struct (see GraphChi documentation (https://github.com/GraphChi/graphchi-cpp/wiki/Introduction-To-GraphChi) for details on GraphChi programs), shown in Supplementary Figure 15. We used the GraphChi scheduler to initialize the program only on vertices belonging to the *VOI.* An *Aggregation Query* is performed in a similar way as for *FlashGraph*, by iterating of the *VOI* vertices and summarizing their outgoing connectivity. In contrast to *FlashGraph*, we summarized them in a vector and not in the receiving vertices, since *GraphChi* loads the edge data everytime a vertex is accessed (i.e. the vertex state and edge lists are not separated), which would lead to loading the edges of the entire graph, just to updated the aggregated connectivity in target/source vertices. Similarly to *FlashGraph*, we forgo to upsample the results to 100-micron standard brain for simplicity.

```
typedef float VertexDataType;
typedef float EdgeDataType;

struct Aggregation_GraphChi_Program : public GraphChiProgram<VertexDataType, EdgeDataType> {
```

```
    std::unordered_set<unsigned> query_vertices; //vertices that belong to VOI
    std::vector<float> resultVector; //vector that contains the aggregated connectivity
    bool loadTargets = false; //if targets or sources should be loaded
    std::mutex mu; //mutex to lock resultVector, since GrpahChiPrograms are executed in parallel

    //Vertex update function.
    void update(graphchi_vertex<VertexDataType, EdgeDataType> &vertex, graphchi_context &gcontext) {
            mu.lock(); //lock
            if(loadTargets){ //if targets or sources should be loaded
                for(int i=0; i < vertex.num_outedges(); i++) {
                    //aggregate connectivity for targets
                    resultVector[vertex.outedge(i)->vertex_id()]+=vertex.outedge(i)->get_data();
                }
            }else{
                for(int i=0; i < vertex.num_inedges(); i++) {
                    //aggregate connectivity for sources
                    resultVector[vertex.inedge(i)->vertex_id()]+=vertex.inedge(i)->get_data();
                }
            }
            mu.unlock();
    }

    //Called before an iteration starts.
    void before_iteration(int iteration, graphchi_context &gcontext) {

        std::vector<float>(gcontext.nvertices).swap(resultVector); //clear resultVector before
iteration
        gcontext.scheduler->new_iteration(1); //this is only to clear tasks in the sheduler!

        //add initialization tasks to sheduler
        for(auto f : query_vertices) {
            gcontext.scheduler->add_task(f, true);
        }

    }

    //Called after an iteration has finished.
    void after_iteration(int iteration, graphchi_context &gcontext) {
    }

    //Called before an execution interval is started.
    void before_exec_interval(vid_t window_st, vid_t window_en, graphchi_context &gcontext) {
    }

    //Called after an execution interval has finished.
    void after_exec_interval(vid_t window_st, vid_t window_en, graphchi_context &gcontext) {
    }
```

Supplementary Figure 15: Aggregation graphchi program struct which is used by the *GraphChi* API to execute *Aggregation Queries*

Supplementary References

Fonov, V., Evans, A. C., Botteron, K., Almli, C. R., McKinstry, R. C., & Collins, D. L. (2011). Unbiased average age-appropriate atlases for pediatric studies. *NeuroImage*, *54*(1), 313–327.

Ganglberger, F., Kaczanowska, J., Penninger, J. M., Hess, A., Bühler, K., & Haubensak, W. (2017). Predicting functional neuroanatomical maps from fusing brain networks with genetic information. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2017.08.070

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., … Jenkinson, M. (2013). The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage*, *80*, 105–124.

Hawrylycz, M. J., Lein, E. S., Guillozet-Bongaarts, A. L., Shen, E. H., Ng, L., Miller, J. A., … Jones, A. R. (2012). An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature*, *489*(7416), 391–399. Retrieved from http://dx.doi.org/10.1038/nature11405

Kyrola, A., Blelloch, G., & Guestrin, C. (2012). GraphChi: Large-Scale Graph Computation on Just a PC Disk-based Graph Computation. *Proceedings of the 10th USENIX Conference on Operating Systems Design and Implementation*. https://doi.org/10.1109/HPCA.2015.7056066

Lein, E. S., Hawrylycz, M. J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., … Jones, A. R. (2007). Genome-wide atlas of gene expression in the adult mouse brain. *Nature*, *445*(7124), 168–176. https://doi.org/10.1038/nature05453

Oh, S. W., Harris, J. A., Ng, L., Winslow, B., Cain, N., Mihalas, S., … Zeng, H. (2014). A mesoscale connectome of the mouse brain. *Nature*, *508*(7495), 207–214. Retrieved from http://www.nature.com/doifinder/10.1038/nature13186

Zheng, D., Mhembere, D., Burns, R., Vogelstein, J., Priebe, C. E., & Szalay, A. S. (2015). FlashGraph: Processing Billion-node Graphs on an Array of Commodity SSDs. In *Proceedings of the 13th USENIX Conference on File and Storage Technologies* (pp. 45–58). Berkeley, CA, USA: USENIX Association.

PAPER C

Iterative exploration of big brain network data

Florian Ganglberger, Nicolas Swoboda, Lisa Frauenstein, Joanna Kaczanowska, Wulf Haubensak, and Katja Bühler. Iterative exploration of big brain network data. In *VCBM 18: Eurographics Workshop on Visual Computing for Biology and Medicine, Granada, Spain, September 20-21, 2018*, pages 77–87, 2018

Eurographics Workshop on Visual Computing for Biology and Medicine (2018)
A. Puig Puig, T. Schultz, and A. Vilanova (Editors)

Iterative Exploration of Big Brain Network Data

Florian Ganglberger[1], Nicolas Swoboda[1], Lisa Frauenstein[1], Joanna Kaczanowska[2], Wulf Haubensak[2] and Katja Bühler[1]

[1]VRVIS Research Center
[2]Research Institute of Molecular Pathology (IMP), Vienna Biocenter (VBC)

Abstract

A current quest in neurscience is the understanding of how genes, structure and behavior relate to one another. In recent years, big brain-initiatives and consortia have created vast resources of publicly available brain data that can be used by neuroscientists for their own research experiments. This includes microscale connectivity data - brain-network graphs with billions of edges - whose analysis for higher order relations in structural or functional neuroanatomy together with genetic data may reveal novel insights into brain functionality. This creates a need for joint exploration of spatial data, such as gene expression patterns, whole brain gene co-expression correlation, structural and functional connectivities together with neuroanatomical parcellations. Current experimental workflows involve time-consuming manual aggregation and extensive graph theoretical analysis of data from different sources, which rarely provide spatial context to operate continuously on different scales.

We propose a web-based framework to explore heterogeneous neurobiological data in an integrated visual analytics workflow. On-demand queries on spatial gene expression and connectivity data enable an interactive dissection of dense network graphs - with of billion-edges on voxel-resolution - in real-time based on their spatial context. In order to take higher order connections between brain regions into account, queries can be executed in a cascading way. Relating data to the hierarchical structure of common anatomical atlases allows experts to quantitatively compare multimodal networks on different scales. Additionally, 3D visualizations have been optimized to accommodate for the domain experts' need for publishable network figures.

We demonstrate the relevance of our approach for neuroscience by exploring social-behavior and memory/learning functional neuroanatomy in mice.

1. Introduction

The quest for understanding the principle organization of the brain and its functional parcellation is constantly changing due to the increasing wealth of multimodal neurobiological data generated by brain initiatives, such as the Allen Institute [all], the Human Brain Project [hum], the WU-Minn Human Connectome Project [VESB*13], and the China Brain Project [mPlDI*16]. They offer rich information about genes, structure and behavior, while understanding their relationship is a key factor in neurocircuit research. Mining these resources can provide researchers with additional context for their experiments. The major challenges for visual analytics in neuroscience involves accessing, fusing and visualizing spatial brain data such as brain-wide gene expression, structural and functional connectivity, and non spatial data like gene lists related to behavior or the functional association of genes.

Our work takes recent advances in circuit neuroscience into account (e.g. neuro- and behavioral genetics, optogenetics, imaging) that identified gene sets underlying a specific behavioral function [KZM*17]. However, there is a lack of tools to explore the mesoscale as well as the global structural and functional brain networks related to these gene sets in silico. We meet this demand by proposing a framework allowing visualization, iterative exploration and the integration of spatial data like imaging data showing brain-wide gene expressions, fMRI, or structural data, with structural, functional and genetic anatomical relations at different scales and different hierarchical anatomical labels. This data can be retrieved from public resources and integrated in our framework. The same applies to private data generated during experiments in the lab.

The data we handle includes hierarchical parcellations and structural annotations (e.g. Allen Mouse Brain and Allen Human Brain region annotations), classical 3D image data (e.g. brain template, spatial gene expression), 3D aggregated data (e.g. Allen Mouse Brain gene expression data or viral injection traces), region-wise connectomes/relations (e.g. resting state functional connectivity) and voxel-resolution connectomes/relations (structural connectivity, gene co-expression), all aligned to a common reference space. Voxel-wise connectivity/relational data generally has a very dense connectivity with up to billions of connections, so their matrices can take up hundreds of gigabytes.

Our framework visualizes volumetric, geometry and graph data simultaneously in 3D rendering and 2D slice views, linked to views showing quantitative profiles at a hierarchical parcellation level.

© 2018 The Author(s)
Eurographics Proceedings © 2018 The Eurographics Association.

DOI: 10.2312/vcbm.20181231

delivered by EUROGRAPHICS DIGITAL LIBRARY www.eg.org diglib.eg.org

The user can interactively navigate the hierarchical levels, which provide spatial context by rendering their brain regions in 2D/3D views. In a previous paper [GKHB18], we described a specialized data structure that organizes and aggregates the voxel-wise connectivity data hierarchically. Via so called *Aggregation Queries* [GKHB18], which represent the aggregated incoming/outgoing connectivity of volumes of interest (*VOI*), the data is made accessible in real time and can be explored interactively on different anatomical scales. The hierarchical organization is anatomy-driven. It can be flexibly generated for different ontologies and their respective spatial region annotations.

In this paper, we used interactive local 3D selection on visualized data to define *VOIs* with high intensity. Based on these, *Aggregation Queries* can be executed on a user selected connectivity data set. The result is the cumulative voxel-wise connectivity of the selected *VOI* that, again is visualized as intensity volume in the 3D rendering as well as quantitative brain-region-wise profile. These are augmented with hierarchical information (i.e. intensity of sub-regions) to provide an overview of different hierarchical levels. This kind of interaction allows the researcher to relate integrated resources, for example incoming/outgoing connectivity at voxel-level and region level, directly to imaging data showing brain-wide gene expressions (spatial gene expression data). Higher order network connections can be targeted by repeatedly cascading over the query results, which can be selected directly in the brain-region-wise profile. This represents an iterative exploration of the networks. The query results can further be visualized as graph representation rendered as 3D and 2D network graphs. This reduces the complexity of voxel-level results to region-level graphs and provides neuroanatomical context, so they can be presented in a way that is particularly suited for neuroscience publications, as advised by domain experts. We integrated the framework components into workflows for interactively fusing volumetric with connectivity data, brain network exploration at different anatomical levels, and intuitive 3D graph visualization. We demonstrate the practical significance of this tool by presenting several use cases based on heterogeneous neurobiological data from large scale brain initiatives that allowed to reproduce several recent biological findings.

Our previous work enabled us to realize, for the first time, methods for interactive exploration of dense voxel-wise brain connectivity data of several gigabytes of size. Based on this, we propose additional visual analytic methods for interactive:

- joint exploration and fusion of brain network data over different scales ranging from multi-resolution voxel-wise connectivity to connectivity defined in respect to different anatomical region hierarchy levels
- joint exploration of data from different brain network types (e.g. structural, functional, gene co-expression)
- identification of higher order connectivity in dense network graphs with billions of edges

2. Background

We have been collaborating with experts interested in neuronal circuits in the mouse brain, how they control emotional states and behavior, and how they are modulated by genes and psychoactive drugs [HKC*10]. The data they use and the related use cases provide a good starting point for understanding how nowadays data-driven research is done in neurocircuit science. Based on previous projects, informal interviews and discussions with our collaborating domain experts we jointly identified the following major analytical workflows and user stories:

Relating spatial gene expression data to different kinds of connectivity: The entry point for many data analytics workflows are 'candidate regions' - brain regions that may be part of a specific brain circuit. Those are affected by genes that are either related to a certain behavior, or targeted by a psychoactive drug. Thus, the knowledge of where a gene affects the brain is a first step in relating it to a particular function. The effect is rather broadly defined. Well documented cases are primary gene expression sites [LHA*07] (sites where the gene creates products, such as proteins) and structural connectivity (regions to which primary gene expression sites project) [OHN*14]. Spatial gene expression and structural connectivity data for the mouse brain is provided e.g. by the Allen Mouse Brain Atlas (*AMBA*) at a voxel-level, that was published by the Allen Institute in 2007 and 2012. This data has been studied for this purpose before [LHA*07]. Here, the primary expression sites of a gene are made available as 8 bit intensity volumes, representing the density of cells expressing the gene (i.e. spatial gene expression data), related to a specific standard brain. To use those resources for hypotheses building, it is necessary to aggregate data manually by querying online databases and literature research. Therefore, an integrated workflow for data fusion is therefore missing. Identifying brain regions that are functionally or structurally connected (either directly or transitively) to this expression site may also contribute to this function or be a second order effect thereof. Performing this tasks at a voxel-level is a particular challenge, since connectivity matrices can easily grow up to hundreds of gigabytes.

Comparing different types of connectivity is of essential importance for identifying neural circuits. For example, two brain regions can have a high structural connectivity (a connection via neurons) but do not necessarily express the same genes (e.g. a so called ligand-receptor binding [YW04]). Therefore, their genetic connectivity for a specific gene set, represented as gene co-expression correlation (the correlation of the gene-expression between two regions) is low. Finding a correlation between regions in their fMRI activity may show their functional relation, but a difference in gene co-expression correlation could reveal that this activity relies on completely different molecular mechanisms. Directional structural connectivity can reveal the flow of information for (undirectional) fMRI connectivity [GSB*15]. The possibilities of brain network comparisons are versatile, but the given examples are in particular relevant for domain experts. Combining specific circuits for larger networks can be done again by manual data aggregation, such as whole brain fMRI studies, but require the expertise of a bioinformatitian.

Exploring the data on different scales: Operating region-wise on the data (for example a region-wise network graph) depends on spatial hierarchy. Brain regions are organized by ontologies as trees with larger regions at the top, and ultimately resolves into voxel-resolution at the lowest level of the related standard brain. In order to keep an overview while comparing global networks

© 2018 The Author(s)
Eurographics Proceedings © 2018 The Eurographics Association.

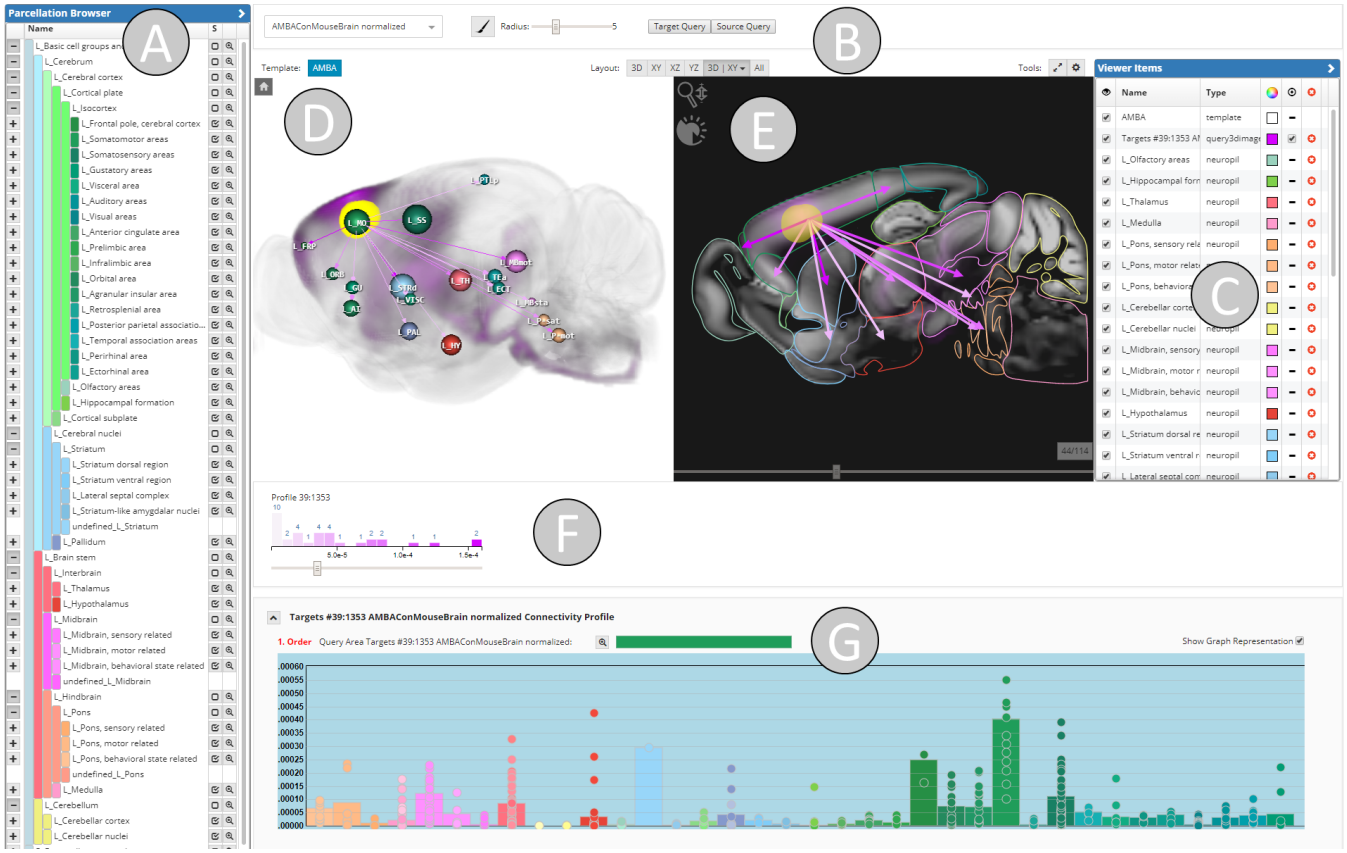


Figure 1: *Web-component for iterative exploration of multiscale brain data.* ***A)*** *Parcellation Browser showing a hierarchical (anatomical) brain parcellation. This controls the parcellation level of visualized graphs and profiles.* ***B)*** *Query Toolbar to start visual queries for target/source connectivity on different connectivity matrices.* ***C)*** *Viewer Items List shows all volumes/geometries visualized in 2D and 3D.* ***D)*** *3D View for rendering volumetric (spatial gene expression and connectivity) data, brain region geometry and network graphs* ***E)*** *2D slice view of 3D representation.* ***F)*** *Filtering Toolbar showing histograms of the visualized graphs and threshold sliders for edge filtering.* ***G)*** *Expression/Connectivity Profile providing a region-wise quantitative representation of volumetric data.*

with different modalities, larger regions are preferred. On the other hand, visualizing small subnetworks for circuit dissection requires smaller regions of the hierarchy or even voxel-level. A hierarchical organization of the network graph can therefore replace time-consuming pre-selection of relevant regions in current experimental workflows.

3. Related Work

Joint exploration of neurobiological data on different scales:

As various studies have already shown for the mammalian brain, combining information of the macroscale connectome with microscale neuronal architecture does provide a deeper understanding of the brain's organization [SSDV14]. Recently, this has also been confirmed for the living human brain, associating resting state functional connectivity with both gene expression [PLK*15] as well as cortical microstructure such as T1-based myelin content, as obtained from ultrahigh-resolution MRI [HBG*17]. This emphasizes the growing need for visualization tools which incorporate data of multiple scales. To visualize the relationship between multiscale data, mentioned studies primarily utilize enhanced heatmaps, as it can be seen in Scholtens et al. [SSDV14], where the matrix entries depict anatomical projections. Structural connectivity strength is encoded by color, projection distance by dot size. To provide spatial context, the matrix is presented side-by-side to a brain surface, color-coding anatomical areas. This approach provides a good overview on region-level. However its complexity increases on higher resolution, such as voxel-level.

To integrate macroscopic data at a brain region level with microscale data from simulation of neural activity, Nowke et al. [NSvA*13] have introduced the interactive analytics tool VisNEST, for Macaque monkeys visual cortex. They provide views for visualizing connectivity between brain regions, within-region connectivity representation and time-varying activities across regions. Regions of interest can be selected in a 3D anatomical view and are represented as meshes, where color and opacity depict activity. Although VisNEST allows comparison of region-level connectivity,

© 2018 The Author(s)
Eurographics Proceedings © 2018 The Eurographics Association.

it is not tailored for high-resolution connectivity data nor does it relate to volumetric data.

For Drosophila, Lin et al. [LTW*11] introduced Neuron Navigator, a tool that allows queries for connectivity of regions of interest in the Drosophila brain space. These spatial queries are achieved by accessing a 3D neuron image database and matching the region of interest with annotated locations of neuron terminals. Although these queries represent connectivity between neurons, they are not executed on actual graph/network data.

BrainExplorer [LNT*08] is an interactive visualization tool, provided by the Allen Brain Institute to explore the spatial gene expression and structural connectivity data provided by the Allen Mouse Brain Atlas. Brain Explorer provides volume rendering and blob-like visualization of gene expressions mapped to a standard brain. Voxels are colored according to their expression level. The tool enables the execution of pre-computed source/target connectivity queries of brain regions, which is also available via a web-interface on the Allen Institute's website [all]. This work comes closest to our solution. However, they provide (pre-computed) queries for incoming/outgoing connections only on pre-defined sites or anatomical regions. Other tools allowing exploration of connectomic data are CATMAID [SCHT09] and ConnectomeExplorer [BAAK*13], which allow tracing neurons on single EM Stacks, so that they work at a local level of a single network with fixed scale. Similar accounts for Sherbondy et al. [SAM*05], who query pre-computed pathways on diffusion tensor imaging data from volumes of interest. NeuroLines [AABS*14] offers a visualization as simplified skeleton graph, similar to a 2D subway map. Different levels of abstraction allow for multi-scale exploration, but is ultimately tailored to branched, tree like connections in electron microscopy data.

A more generic framework, named neuroVIISAS [SE12] provides multiple ways of organizing, visualizing and analyzing multi-scale brain network simulation data. NeuroVIISAS offers a collection of visualization techniques in 2D such as different representations of hierarchical connectivity matrices, circular connectograms, various layouts for planar graphs, as well as in 3D, where regions are rendered as spheres or true anatomical surface meshes, colored based on the color defined in the reference ontology. Connections are depicted as tubes with or without arrows. Moreover, individual source and target queries can be performed textually by defining filters on a table of all the available connections. The resulting selections can be viewed side-by-side in 2D atlas views. However, as this tool is designed for analysis of simulation data it has a different scope.

Visualization of structural, functional and gene co-expression brain networks: Brain networks are commonly represented as node-link diagrams, connectivity matrices or connectograms, where only the first can be visualized in anatomical space both in 2D, with slice views of the anatomical planes and 3D. In volumetric space, edges have been rendered as straight lines in between nodes depicted as spheres across the brain [XWH13, LDTS14] or curved lines along the brain surface [LFG*15]. However, with a dense number of connections, this produces clutter and obscures the underlying anatomy. Mostly, brain networks are too complex to be able to show all details up to the highest level of resolution whilst seeing the entire global structure. Thus, hierarchies can define the level of partition. Applied to rendering in 3D space, the hierarchical level can be determined by the distance to the viewpoint in a level-of-detail-visualization [BD07], edge bundling [BSL*14]. Hence, brain network visualization tools [MBB*16] often present network structure and anatomical context side-by-side and achieve reference through linked views and coloring network nodes according to their anatomical region. This provides the flexibility to layout the connectivity network according to network properties such as using a spring-embedded layout where well-connected groups of nodes are pulled together as seen in [PLK*15]. Different graph layouts, such as an anatomical layout, are also supported by NeuroMap [Sor13], which renders potential neuronal connections in the fruit fly's brain as interactive circuit-style wiring diagrams. Spatial context is provided by introducing brain regions as spatial constraints to a 2D layout by providing a 2D abstraction of anatomical organizations and by linking NeuroMap with 3D visualizations, showing the neurons in their 3D anatomical context. Although these tools provide spatial/anatomical context, they are limited to a regional level, and do not scale to voxel-level resolution with billions of edges.

4. Core Tasks

Based on the user stories described in Section 2 we identified the following core tasks and requirements to be supported by our system. It is necessary that all tasks can be performed interactively i.e. without time-consuming computation:

R1: 3D target/source assessment: *What is connected to a certain volume of interest?* The user has 3D volume data that shows activity in certain areas, for example where a certain gene is expressed. She wants to know which areas are connected via structure or by function at a voxel-level. She can directly select the areas that are interesting for her based on the visualization and receives the results instantly. The user can select additional connectivity modalities, which she can visualize similarly or by their overlap.

R2: Higher order target/source assessment: *What is transitively connected to a certain volume of interest?* Originating from the target/sources of R1, the user wants to know what further connections exist i.e. which brain regions (also on a voxel-level) are transitively connected to a certain area. Using R1 iteratively is simply too time consuming for the user, since it involves the accurate selection of potentially large brain areas.

R3: Anatomical context: *To which brain regions is my data related?* 3D volume data, such as spatial gene expression data or target/source connectivity query results need to be related to their anatomical context, so that the user has quantified information of what she is seeing. Therefore, the volume data needs to be summarized to show how much gene expression/connectivity individual brain regions exhibit. The hierarchical nature of the regions makes it necessary to navigate/visualize different hierarchical levels.

R4: Explore 3D network graphs with spatial reference: *How do my networks look like on different scales?* The user wants to explore connectivity of experimental data, such as a fMRI network of a study. She executes target/source queries on the fMRI network and

© 2018 The Author(s)
Eurographics Proceedings © 2018 The Eurographics Association.

other modalities and renders a graph representation of the results. She wants to navigate hierarchical brain parcellation until the graph fits her needs. Multiple connectivities need to be represented as a graph for purposes of comparison, i.e. to see the difference from and the overlap with her experiment.

R5: Publishable figures: *Give me a publishable figure!* The region-wise network graph, generated by R4 needs to be visualized in a way with which neuroscientists feel intuitively familiar. The figure needs to be easy to understand, interpretable and suitable for neuroscientific publications/presentations. The domain experts consider the best solution to be a 3D graph visualization overlayed by the outlines of the brain. Furthermore, the workflow for generating the figure has to be traced in order to be reproducible.

5. A Data Structure for real-time Aggregation Queries of Big Brain Networks

Querying large graphs, with billions of edges can be time consuming. This data, represented as connectivity matrices (weighted adjacency matrices) can easily grow up to hundreds of gigabytes, which is why keeping them in memory becomes infeasible with increasing size. This requires special solutions for fast disk access (R1, R2). When operating on different anatomical scales, it is necessary to perform cumulative operations on the connectivity matrices (e.g calculate region-level connectivity from voxel-wise connectivity). In this case large parts of the connectivity matrix need to be loaded and processed. We created a data structure in our previous work [GKHB18] to allow these *Aggregation Queries* to happen in real-time, directly on the hard disk. It uses a specialized *Connectivity Storage* to efficiently manage and access large connectivity data (Figure 2). By exploiting the sparseness of the data and its spatial organization, it optimizes disk-reading speed via read-ahead paging and therefore achieves almost sequential reading-speed for local *Aggregation Queries* (i.e. cumulated connectivity of a brain area). To scale for larger brain areas, we implemented a cache mechanisms that uses pre-computed queries in a hierarchical way, similar to an image pyramid. Moreover, the *Connectivity Storage* provides a mapping to a standard reference space, which allows the retrieval of connectivity data regardless of its original resolution. The *Connectivity Storage* has been combined with a multi-model graph/document-database, which is tailored to hierarchical brain parcellations (i.e. hierarchical anatomical brain atlases) to store region-level connectivity for different hierarchy levels. It is also used to organize spatial data, such as volume data (e.g. spatial gene expression), binary masks and meshes of region definitions.

A REST API is acting as the central access point for the data. It provides calls for querying the connectivity matrices, as well as importing them. Data import can be performed by providing the matrices in a row/column table format as csv. The data structure will then convert the data to a compressed file format for the *Connectivity Storage* and automatically aggregats region-level connectivity in the graph-database (see details in [GKHB18]). It can then be used immediately afterwards for real-time connectivity queries. These can be executed in less than one second for brain areas involving about 1% of the brain, and less than four seconds for larger areas up to the whole brain on SSD. On HDD the queries are approximately three times slower.

6. Visualization Components

6.1. Parcellation Browser

The *Parcellation Browser* (Figure 1A) shows a hierarchically organized brain parcellation in a tree view, where every brain region has a name and a color code. Its main purpose is to define the brain regions that are visualized in the *Expression and Connectivity Profiles* as well as the query result's graph representation (the *selection state*). Navigating the tree adds/removes regions to/from the *selection state* to let the user control the current level of detail (first column), or mask parts of the brain (column *S*), depending on her research question. The uppermost regions are the two hemispheres that can be expanded further down the hierarchy. Each major brain region has a color code that is varied in shade depending on its hierarchical level, providing the user with an indirect reference to the larger anatomical context of a subnode and to the spatial position in the brain. This effect is further enhanced by using established color codes for brain regions together with the brain ontologies, which, for example, available in the *AMBA* context.

6.2. Query Toolbar

The *Query Toolbar* (Figure 1B) can be used to execute target/source queries on connectivity matrices (R1), selected by a dropdown menu. Furthermore, it can start brushes, which is a spherical selection tool in the 2D slice view.

6.3. Viewer Item List

The *Viewer Item List* (Figure 1C) controls the visibility and appearance of spatial volumetric data and geometry that is visible in the application. Visibility of volumetric data can be controlled by setting brightness, contrast and transparency. Moreover, it contains entries for all regions in the selection state, since they are visible in 2D (color-coded in the same way as in the *Parcellation Browser*. For all items, the visibility and color, and for query results (i.e. connectivity data), the visibility of the graph representation can be set.

6.4. 3D Visualization in Anatomical Space

To visualize data in its original spatial, anatomical environment we are using classical volume rendering (Figure 1D) and multi-planar reformation (Figure 1E) for the visualization of volumetric data and graph rendering in a ball-and-stick model, since they are common in the neuroscience community. We further render anatomical context in the planar view, visualized by region contours, and mesh geometry in 3D (if enabled via *Viewer Items List*) in order to provide anatomical context. 3D mesh rendering is turned off by default, since it would obstruct 3D volume rendering.

6.4.1. Visualization of Volumetric Data

Volumetric data, like gene expression data or voxel-level structural connectivity is rendered in 3D (Figure 1D), as well as three planar slice views (Figure 1E). A maximum of four volumes can be blended simultaneously in 3D using single color-based transfer-functions (limited by the four RGBA channels, since the volume is loaded as texture). In the slice views, blending is also used for multiple images. Therefore, the rendering of intensity-overlap converges to white.

© 2018 The Author(s)
Eurographics Proceedings © 2018 The Eurographics Association.

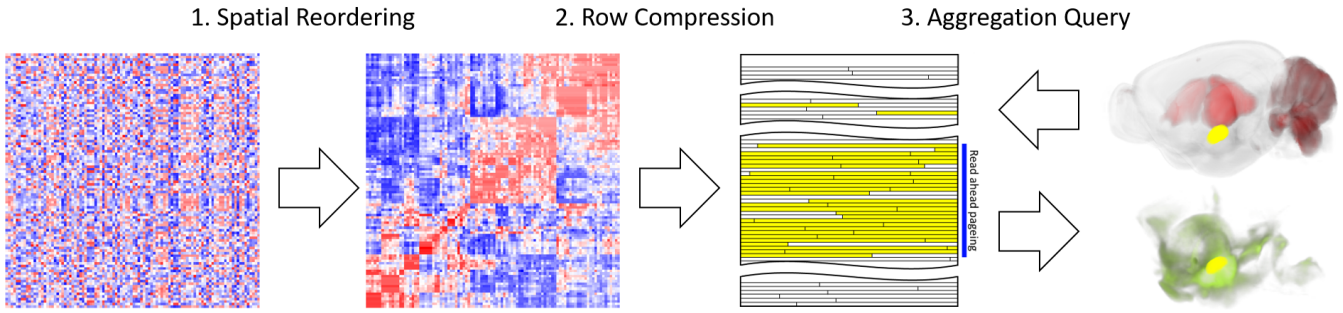


Figure 2: *Principle of the Connectivity Storage and real time Aggregation Queries:* ***1.*** *Spatial reordering of a (voxel-level) connectivity matrix. Rows/columns that represent the outgoing/incoming connectivity of voxels are reordered, so that rows/columns that belong to spatially close voxel are close to each other.* ***2.*** *Row Compression: The reordered connectivity matrix is stored with a row-wise compression as Connectivity Storage File on the hard disk. Row-wise compression saves disk space (and therefore reading time) by exploiting sparseness.* ***3.*** *Aggregation Query: A volume of interest (yellow) is selected on 3D spatial data (e.g. spatial gene expression data, red). Outgoing connectivity is aggregated by reading the corresponding rows (yellow) from the Connectivity Storage File. Since spatially close rows are also close in the file, they can profit from read ahead paging (blue). The result is the aggregated connectivity for every voxel (green).*

6.4.2. Region-level Graph Representation of Voxel-Wise Connectivity

A result of a target/source query can be also visualized as a 3D ball-and-stick-model within its anatomical reference space (R4 and R5). Network nodes are rendered as spheres, where size is determined by the size of the corresponding anatomical region and color reflects their anatomical representation in the originating parcellation. Rendering them as meshes instead, would cause blocking the view by larger brain regions. For directed networks, edges are rendered as tubes with arrows, differentiating incoming from outgoing connections. Undirected network edges are represented by tubes. For weighted networks, connectivity strength is color-coded, ranging from white to the color selected in the *Viewer Item List*. Rendering arrows that are indicating the direction of connections, were specifically requested by our expert users as they are an intuitive way of differentiating incoming from outgoing edges in publications. Edges and nodes are user-adjustable via color-transfer-functions. Moreover, thresholding edge weights enables dynamical network refinement and an exploration of connections of interest. Node labels further support orientation. In addition to the network model in the 3D view, network regions are shown as colored outlines of their anatomical equivalent in the corresponding 2D slice views; arrows (for directed) or lines (undirected networks) are rendered between the region centers in every slice. Although this shows only a part of the network (i.e. in the specific slice), it has the advantage that one can corectly render the regions by anatomy (instead of using a sphere), without facing the problem caused by possible obstruction in 3D.

6.4.3. Filtering Connectivity Strength

Since the mean number of edges increases quadratically with the number of nodes, the user can apply thresholding on edges to highlight stronger connections. This is done with a slider to set a user defined threshold, in combination with a histogram that shows the edge weight distribution in the *Filtering Toolbar* (Figure 1F). The colors of the histogram bars map directly to the visualized edges. Hence, color directly relates to connectivity strength.

6.4.4. Multiple Network Graphs

The comparison and joint exploration of multiple graphs, which relate to the same anatomical parcellation, has been realized by overlaying several graphs, i.e. simply rendering multiple edges between nodes (Figure 4E). Our domain experts found this the most intuitive method, since one can directly see the similarity/contrast of two edges between two nodes. Showing only two graphs even renders up to four arrows between nodes. Therefore, we also implemented an overlap visualization. Between two nodes, a maximum of two edges, whose weight is defined by the multiplication of all edges (at region-level) between those nodes, is rendered. If an overlap is computed, the histograms depict a formula of the multiplication of individual graphs (Figure 4F), so that the calculation can be comprehended by the user. Cascading queries are treated as a single network, depicted by brackets in the formula (see Section 7.2), since they represent connectivity of different orders.

6.5. Connectivity and Expression Profiles

Connectivity Profiles (*C-Profiles*) as well as *Expression Profiles* (*E-Profiles*) (Figure 1G) summarize volumetric voxel-level data in order to make them quantitatively comprehensible for the user (R3). They visualize the mean target/source connectivity or gene expression for brain-regions in bar graphs. These brain regions are defined by the selection made in the *Parcellation Browser*. The mean connectivity/expression of all subregions are rendered as dots. This allows the user both to identify regions that have highly connected/expressed subregions, and to further refine the anatomical hierarchy to focus on relevant subregions. For anatomical context, the colors of the bars and dots are corresponding to the *Parcellation Browser*. Names are made visible via a tooltip. One or multiple regions can be picked for a high-intensity *VOI* selection, which

© 2018 The Author(s)
Eurographics Proceedings © 2018 The Eurographics Association.

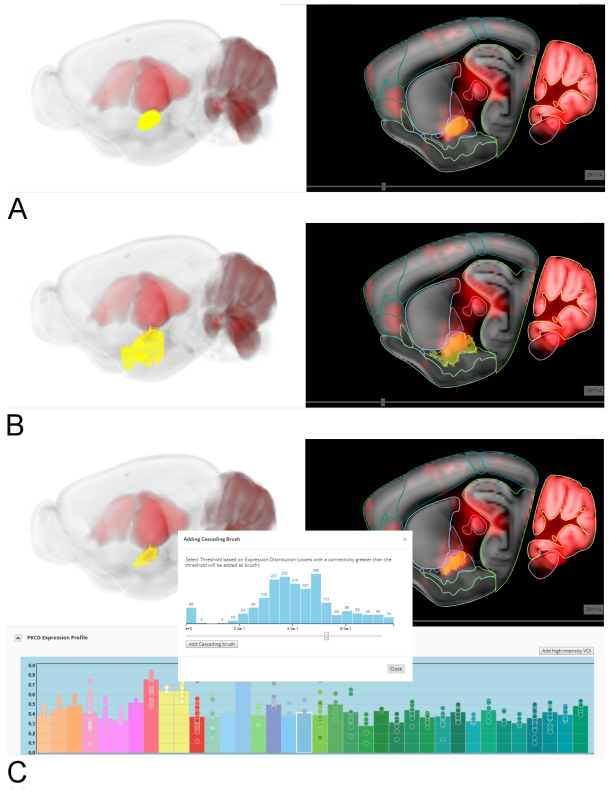


Figure 3: *Selection of parts of a gene expression pattern (red). **A)** Brush-selection of spherical region (yellow) in 2D slice views and simultaneous rendering in 3D. **B)** Region-selection initiated by the Parcellation Browser, voxel-level visualization of a region (yellow). **C)** High-intensity-selection of a brain-region started from an Expression Profile. A dialog which shows a histogram of voxel-level connectivity within the selected region lets the user decide which intensity, and therefore which voxels, will be selected. The result is rendered in 2D and 3D (yellow).*

enables the user to threshold for high expression/connectivity voxels in a separate dialog (see Figure 3). This allows higher order target/source queries in an iterative workflow (R2), as further described in Section 7.1.3. For *C-Profiles*, the percentual composition of the query region is shown in a bar above the profile to indicate the queries anatomical context. It uses the same parcellation/color code as the profile, and reveals its region names via mouse-over. Furthermore the query region can be added again to the 2D/3D viewer, so that a user can reproduce the query with the same or other connectivity matrices.

7. Basic Workflows for Joint Data Exploration

7.1. Visual Queries

A visual query allows API requests to the data structure that are based on selections of a volume of interest directly in volume rendering. The response can then be directly shown in the 2D/3D renderer as well as *C-Profile*. This subsection describes the interaction with volume data, and the query types we have implemented.

7.1.1. Selecting a Volume of Interest

In the volume rendering, the selection of areas can be performed in three different ways: Brush-selection, region-selection, and high-intensity-selection. Brush-selection is performed from the *Query Toolbar*. It lets the user draw pherical areas encoded in transparent yellow in the 2D slice views, which are also directly rendered in 3D. Figure 3A for example shows a gene-expression volume, where the spherical area is drawn on voxels with high gene-expression. Executing queries on this area is therefore acting as link between volume and connectivity data. The region selection (Figure 3B), added by the magnifying-glass-plus-button in the *Parcellation Browser* next to a region, selects only voxels within this region rendered as transparent yellow cubes. This provides the user with the possibility to explore the network without volume data. High-intensity queries are started from the *E-Profiles*. It allows the user to select voxels with high-intensity within user-selected brain regions. The user selects brain regions of interest in the *E-Profiles* (Figure 3 C) with high gene expression. By clicking the "add high-intensity VOI " button, a dialog appears which shows a voxel-level histogram of the intensity values within this region. Here the user defines high-intensity voxels by setting a threshold. In the 3D/2D views, the voxels are instantly selected and visualized similarly to the region selection.

7.1.2. Target/source queries

Target/source queries can be used to link connectivity data with volume data (R1). In the *Query Toolbar*, connectivity matrices for querying can be selected and brush-drawings with a certain radius can be started (although one can also use the other brush-types instead). The selected area (Figure 3A) then acts as an input for a target or source query on the API. The API retrieves the connectivity to all voxels that are either targets or sources from the selection, and then returns a *Connectivity Volume* (connectivity from/to VOI on voxel-level) as compressed JPEG to the web-component which will be rendered instantly in 3D, 2D slice view and as *C-Profile* (R3). This represents the cumulated connectivity to (target) or from (source) the selected area. The *Connectivity Volume* automatically gets assigned a random colorscale, from white to a color that differs the most in its RGB value from other viewer items, to make it visually most distinctive to other visualized volumetric data. *C-Profiles* are shown in a list under each another, with the most recent one on top, so that the user does not need to scroll down.

7.1.3. Higher order target/sources

On the *C-Profile* of the *Connectivity Volume*, the user can select regions of interest with high connectivity to start a high-intensity-selection (Figure 3C). Similarly to the *E-Profile* (Section 7.1.1), the user can choose a threshold in a voxel-level connectivity histogram, which adds all voxel within the region, and a connectivity above the threshold, as *VOI*. After performing a target/source query on the selection, the resulting 2nd-order connectivity is visualized in 2D and 3D as other volume data, while the associated *C-Profile* will be rendered below its originating *C-Profile* (R2). The name is indented

© 2018 The Author(s)
Eurographics Proceedings © 2018 The Eurographics Association.

and prefixed with "2nd-order" in red font, so that it is clear to the user that this query belongs to an iterative cascade, while the profile itself is not indented to ensure vertical comparability between multiple profiles (Figures 5 C). The cascading nature of these queries is further emphasized in the region-level graph representation, described in the next section.

7.2. Transformation of Voxel-level Connectivity to Region-level Graph Representation

Changing from voxel-level *Connectivity Volumes* to region-level graph representation can be controlled either directly from a *C-Profile* via "Show Graph Representation" button, or via the *Viewer Item List* (R4). The *Graph Representation* is a mapping of already aggregated connectivity to region-level, and therefore does not reflect the total and individual connectivity between brain regions visible in the graph. It rather represents an abstract view of the query, showing the aggregated connectivity of a *VOI* mapped to a region-level in its 3D/2D spatial context. Moreover, it allows iterative queries for higher order target/sources to be visualized in a cascading way, since the individual graphs are connected by arrows (Figure 5D) (R5). Edges of the graph are drawn from all regions in the query area, to all regions in the *C-Profile* (i.e. the selection state), with their mean connectivity as weights. These edges do not need to be calculated, since they have already been aggregated for the *C-Profiles*.

7.3. Region-level Selection and Manipulation

The central control element for the *C-Profile* and *Graph Representation* is the *Parcellation Browser*. Navigating in the tree directly influences the selection state, thereby determining which bars (profile) or nodes (graph) are shown (R3). Opening a subtree of a region will remove the corresponding bars/nodes from the graph, and adding all its children. Closing a region will add it, while removing its children. Since the user is not limited by a rigid parcellation, she can focus on regions of the brain that are relevant to her (e.g. only sub-cortical areas),

8. Case Studies

We conducted two case studies, which were designed in colaboration with our domain experts. The first case study compares different kinds of networks (structural-connectivity vs gene co-expression correlation) related to social-bonding behavior. The second examines neurocircuits related to memory and learning, which we explore iteratively.

8.1. Comparison of multiple networks

In the first case study the domain experts would like to examine brain networks related to social-bonding behavior. In particular, they were interested in oxytocin and vasopressin (neuropeptides known to be related to social behavior in mammals) release effect at the network level [BA15]. For this, they wanted both to examine primary expression sites of both genes, and explore their target sites (i.e. outgoing connectivity of expression sites) on different networks. For this case study we incorporated the following: spatial

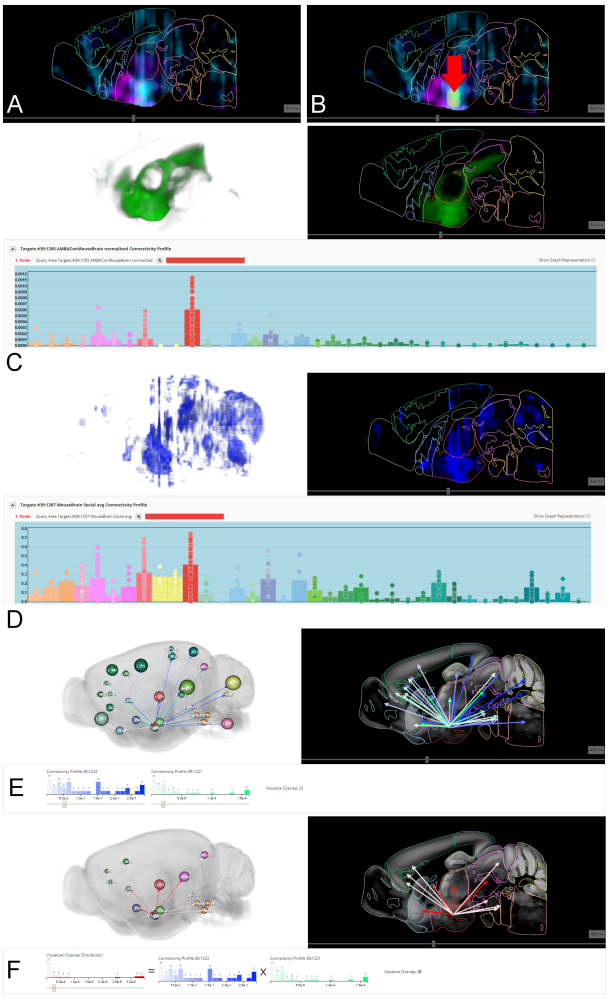


Figure 4: *Case Study 1:* ***A)*** *Gene expression of OXT (cyan) and AVP (purple).* ***B)*** *Selected VOI with brush-selection (yellow, indicated by red-arrow).* ***C)*** *Structural connectivity of the VOI in 3D and 2D (green) and its Connectivity Profile.* ***D)*** *Gene co-expression correlation of the VOI in 3D and 2D (blue) and its Connectivity Profile.* ***E)*** *Graph representation of structural connectivity (white to green) and gene co-expression correlation (white to blue).* ***F)*** *Multiplication (overlap) of the connectivities (red).*

gene expression data (67x41x58 volume on a 200-micron resolution) of oxytocin (OXT) and vasopressin (AVP), a spatial gene co-expression correlation network (also 200-micron resolution, matrix file size is $\sim$ 12 GB) of social-bonding related genes consisting of (gene ENTREZ ID in brackets) Oxt (18429), Oxtr (18430), Avp (11998), V1b receptor (26361), D1R (13488), D2R (13489), Slc6a3 (13162) and Crh (12918); and a 100-micron voxel-level (132x80x114) structural-connectivity from the *AMBA* (file size is $\sim$ 90 GB) [GKHB18]. Although these networks have billions of edges, our data structure [GKHB18] allows real-time retrieval of aggregated connectivity (i.e. cumulated incoming/outgoing connectivity) of a *VOI* on a 132x80x114 standard brain space.

© 2018 The Author(s)
Eurographics Proceedings © 2018 The Eurographics Association.

Identify a volume of interest of spatial data: The entry point for experts is the spatial gene expression data of OXT (Figure 4A cyan) and AVP (Figure 4A purple). The user notices an overlap in the hypothalamus (indicated by dark red contours, in the 2D slice view) (R3). She selects the overlapping area with a spherical brush initiated by the *Query Toolbar* (Figure 4B yellow spheres, highlighted by red arrow).

Find connected areas: The user selects the structural connectivity matrix in the *Query Toolbar*, a click on the *Target Query* button executes the selected *VOI* on the data structure (R1). The accumulated connectivity is instantly displayed as *Connectivity Volume* in 3D and 2D (Figure 4C green) and as *C-Profile* (Figure 4C bottom). The profile shows that the strongest connectivity is to the Hypothalamus (dark red) itself (strongest connectivity to itself is common for structural and gene co-expression correlation). Other top connections are the Striatum (light blue), Pallidum (dark blue), Midbrain (pink), Thalamus (light red) and some cortical areas (green) (R3). This process is repeated for the gene co-expression correlation (Figure 4D blue), where its profile depicts hypothalamus as strongest again, but among similar connections also strong connections to Cerbellum (yellow) and Hindbrain (orange).

Compare networks: To compare the connectivities, the user visualizes them as *Graph Representation* (R4). Figure 4E shows the two graphs in 2D and 3D, easily recognizable by their colors. In the *Filtering Toolbar* the user selects a threshold to filter weak connections. Among others, the user sees that both networks show projections to the Septal Complex (LSX), the Anterior Cingulate (ACA) and the Prelimbic/Infralimbic Area (PL and ILA), which are well known nodes involved in social behavior [BA15]. To further highlight this, the user also visualizes the overlap of both graphs (i.e. multiplication of both connectivities). The resulting reduced amount of edges makes it easier for the user to identify regions which are connected via both connectivities. These types of network visualization were familiar to the domain experts, who confirmed their suitability for neuroscientific visualization or publications (R5).

This case study illustrates the joint exploration of gene expression data and different kinds of connectivities on voxel- and region-level. This approach allows real-time visual analytic workflows which are fast and efficient, compared to time-consuming manual data aggregation by querying different online databases, literature research and scripting. A video of the case study is available as *Supplementary Video 1*.

8.2. Higher-order connectivity

We designed this case study with domain experts to showcase iterative higher-order connectivity queries with a well-known relationship between the brain-derived neurotropic factor (BDNF) and hippocampal synaptic plasticity, respectively circuits related to learning and memory [LBD17]. One of these circuits is Dentate Gyrus (DG) → CA2/CA3 → CA1 → Entorhinal cortex/Subiculum [LBD17]. For this purpose, we explored a primary expression site of BDNF in DG, and traversed their first-, second- and third-order targets iteratively. This case-study required the spatial gene expression data at a 200-micron resolution for BDNF (gene ENTREZ ID: 12064) and structural connectivity at a 100-micron resolution from *AMBA* in the data structure [GKHB18].

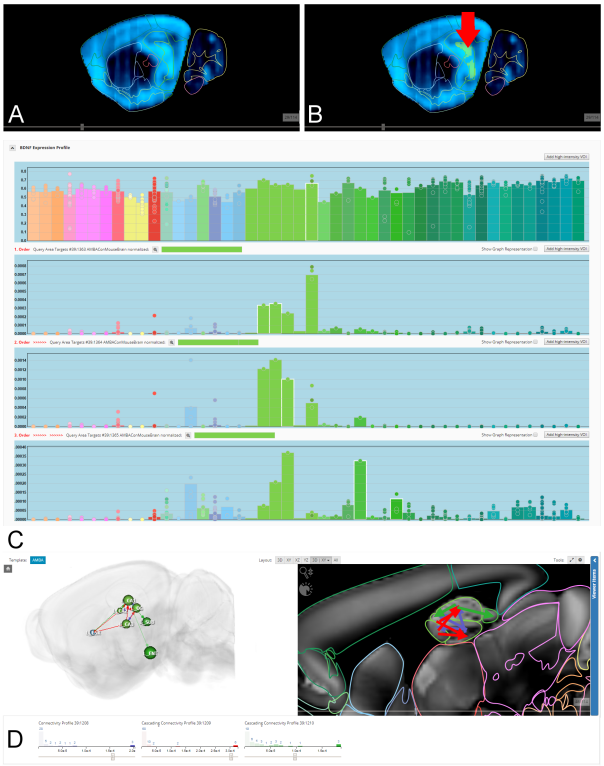


Figure 5: *Case Study 2:* ***A)*** *Gene expression of BDNF (light blue).* ***B)*** *Selected VOI with high-intensity-selection (yellow, indicated by red-arrow).* ***C)*** *Expression Profile of BDNF and its first-, second- and third.order Connectivity Profiles* ***D)*** *Graph representation of the connectivity in 2D and 3D (first-order blue, second-order red, third-order green).*

Identify a volume of interest of spatial data: The experts started their investigation by visualizing the spatial gene expression of BDNF in 2D/3D and as *E-Profile* (Figure 5A). Since the circuit has its origin in DG, the user navigates in the *Parcellation Browser*. In the 2D slice view the user sees that there is a high expression level, hence she starts a high-intensity-selection in the *E-Profile* (Figure 5B).

Find connected areas iteratively: By selecting the structural connectivity matrix in the *Query Toolbar* and clicking the *Target Query* button, the user receives the accumulated connectivity instantly as 2D/3D visualization and *C-Profile*. The name of the profile is indented and prefixed with "First-Order" to highlight the iterative procedure (Figure 5C). CA2 and CA3 are the strongest connections (not counting DG, since it is a connection to itself), so they are chosen for the next high-intensity-selection to go further along the circuit (R2). Note that CA1 receives strong input, but this is primarily caused by their spatial closeness and data acquisition technique for structural connectivity [AMB]. The next *C-Profile* shows CA1 as strongest connection (except for the originating CA2/CA3).

© 2018 The Author(s)
Eurographics Proceedings © 2018 The Eurographics Association.

By performing a high-intensity-selection on CA1, the results show strong connectivity in the Retrohippocampal Region. After browsing its subregions in the *Parcellation Browser*, the connections to Entorhinal Cortex and Subiculum are revealed (highlighted in Figure 5C bottom)

Visualize the Circuit: Finally, the experts wanted to see the circuit in a 2D-3D *Graph Representation* to give them a spatial context (R4), so that they can use their exploration for presentation and discussion purposes with colleagues (R5). They selected the "Show graph representation" button next to the profiles, which instantly shows them the network graphs (Figure 5D). After filtering for the strongest connections, the graph shows only the DG → CA2/CA3 → CA1 → Entorhinal cortex/Subiculum circuit (and a connection to the lateral septal complex LTX, which is not relevant for this case study). Since the *C-Profiles* were automatically marked with different colors, the colored histograms in the *Filtering Toolbar*, that are ordered according to their iteration, allow association of graph edges with the connectivity-order originating from DG.

In this case study, we showed the iterative exploration of a 90 GB connectivity matrix at voxel-level, enabled by quantitative information in *C-Profile* and high-intensity *VOI* selection. To our best knowledge, this could be done so far only at region-level connectivity. We provide a video of this study as *Supplementary Video 2*.

9. Discussion and Conclusion

In this paper we present a novel integrated workflow for analyzing and fusing heterogeneous neurobiological data of different types, modalities and scale via their spatial context. The workflow is integrated in a framework that combines data from different large-scale brain initiatives with user generated data. We incorporate a hierarchically organized data structure, which enables real-time querying and aggregating of huge brain network connectivity of different scales and resolutions on a common standard brain space. This data structure can be accessed via a web-component, which allows selecting *VOI* on 2D/3D visualizations of various volumetric data. The resulting voxel-level connectivity is rendered again in 2D/3D, where it can directly be compared to spatial gene expression or connectivity of different modality. Importantly, these operations remain interactive, despite operating on matrices with up to hundreds of gigabytes.

To allow a quantitative evaluation on brain-region level, the results are shown as region-wise profile, presenting the mean-connectivity for every brain region, as well as for all its sub-regions in a bar chart. This directly highlights strongly connected subregions, which would otherwise be missed if the user operated at a higher brain region level. The region-level profile of connectivity further allows for *VOI* selection for transitive connectivity. Although it would generally be possible for a user to manually select strongly connected voxels in the 2D slice view in order to start another target/source query, the domain experts preferred to query for transitive connectivity within brain regions of their interest (i.e. they were not interested in retrieving the connectivity of all target/source sites, but only from sites within brain regions of interest). Therefore, the user selects strongly connected brain-regions in the region-level profile, chooses a connectivity threshold and receives a voxel-level selection that can be used for further queries. Query results can be abstracted as 2D and 3D network graphs, which reduces the complexity of voxel-level while incorporating neuroanatomical context. Multiple graphs (i.e. region-level graph representation of connectivity query results) can be rendered with parallel arrows/lines, or combined by multiplication. Domain experts confirmed that they feel familiar with this type of visualization. Moreover, it is suitable to be used for discussion or as figures in neuroscientific publications.

Finally, the case studies conducted with domain experts showed biological validity and reproducibility by reproducing findings of known microcircuits that are subject to current research.

10. Acknowledgments

This work is a result of a joint IMP VRVis project supported by Grant 852936 of the Austrian FFG Funding Agency. W. H. was supported by a grant from the European Community's Seventh Framwork Programme (FP/2007-2013) / ERC grant agreement no. 311701, the Research Institute of Molecular Pathology (IMP), Boehringer Ingelheim and the Austrian Research Promotion Agency (FFG). VRVis is funded by BMVIT, BMDW, Styria, SFG and Vienna Business Agency in the scope of COMET - Competence Centers for Excellent Technologies (854174) which is managed by FFG.

References

[AABS*14] AL-AWAMI A. K., BEYER J., STROBELT H., KASTHURI N., LICHTMAN J. W., PFISTER H., HADWIGER M.: Neurolines: A subway map metaphor for visualizing nanoscale neuronal connectivity. *IEEE Transactions on Visualization and Computer Graphics 20*, 12 (Dec 2014), 2369–2378. doi:10.1109/TVCG.2014.2346312. 4

[all] Allen institute. URL: https://www.alleninstitute.org/. 1, 4

[AMB] Allen mouse brain atlas whitepapers. URL: http://help.brain-map.org/display/mouseconnectivity/Documentation. 9

[BA15] BARIBEAU D. A., ANAGNOSTOU E.: Oxytocin and vasopressin: Linking pituitary neuropeptides and their receptors to social neurocircuits, 2015. doi:10.3389/fnins.2015.00335. 8, 9

[BAAK*13] BEYER J., AL-AWAMI A., KASTHURI N., LICHTMAN J., PFISTER H., HADWIGER M.: Connectomeexplorer: Query-guided visual analysis of large volumetric neuroscience data. *IEEE Transactions on Visualization and Computer Graphics (Proceedings IEEE SciVis 2013) 19*, 12 (2013), 2868–2877. 4

[BD07] BALZER M., DEUSSEN O.: Level-of-detail visualization of clustered graph layouts. *Asia-Pacific Symposium on Visualisation 2007, Proceedings* (2007), 133–140. 4

[BSL*14] BOETTGER J., SCHAEFER A., LOHMANN G., VILLRINGER A., MARGULIES D. S.: Three-Dimensional Mean-Shift Edge Bundling for the Visualization of Functional Connectivity in the Brain. *IEEE Transactions on Visualization and Computer Graphics 20*, 3 (2014), 471–480. 4

[GKHB18] GANGLBERGER F., KACZANOWSKA J., HAUBENSAK W., BUEHLER K.: A data structure for real-time aggregation queries of big brain networks. *bioRxiv* (2018). doi:10.1101/346338. 2, 5, 8, 9

[GSB*15] GENÇ E., SCHÖLVINCK M. L., BERGMANN J., SINGER W., KOHLER A.: Functional connectivity patterns of visual cortex reflect its anatomical organization. *Cerebral Cortex* (2015), bhv175. 2

© 2018 The Author(s)
Eurographics Proceedings © 2018 The Eurographics Association.

[HBG*17] HUNTENBURG J. M., BAZIN P.-L., GOULAS A., TARDIF C. L., VILLRINGER A., MARGULIES D. S.: A Systematic Relationship Between Functional Connectivity and Intracortical Myelin in the Human Cerebral Cortex. *Cerebral Cortex* (2017), 1–17. 3

[HKC*10] HAUBENSAK W., KUNWAR P. S., CAI H., CIOCCHI S., WALL N. R., PONNUSAMY R., BIAG J., DONG H.-W., DEISSEROTH K., CALLAWAY E. M., ET AL.: Genetic dissection of an amygdala microcircuit that gates conditioned fear. *Nature 468*, 7321 (2010), 270–276. 2

[hum] Human brain project. URL: https://www.humanbrainproject.eu. 1

[KZM*17] KIM J., ZHANG X., MURALIDHAR S., LEBLANC S. A., TONEGAWA S.: Basolateral to central amygdala neural circuits for appetitive behaviors. *Neuron 93*, 6 (2017), 1464–1479. 1

[LBD17] LEAL G., BRAMHAM C., DUARTE C.: Chapter eight - bdnf and hippocampal synaptic plasticity. In *Neurotrophins*, Litwack G., (Ed.), vol. 104 of *Vitamins and Hormones*. Academic Press, 2017, pp. 153 – 195. doi:https://doi.org/10.1016/bs.vh.2016.10.004. 9

[LDTS14] LAPLANTE R. A., DOUW L., TANG W., STUFFLEBEAM S. M.: The Connectome Visualization Utility: Software for Visualization of Human Brain Networks. *PLoS ONE 9*, 12 (2014). 4

[LFG*15] LI H., FANG S., GONI J., CONTRERAS J. A., LIANG Y.: Integrated Visualization of Human Brain Connectome Data. *Lecture Notes in Computer 9250* (2015), 295–305. 4

[LHA*07] LEIN E. S., HAWRYLYCZ M. J., AO N., AYRES M., BENSINGER A., BERNARD A., BOE A. F., BOGUSKI M. S., BROCKWAY K. S., BYRNES E. J., ET AL.: Genome-wide atlas of gene expression in the adult mouse brain. *Nature 445*, 7124 (2007), 168–176. 2

[LNT*08] LAU C., NG L., THOMPSON C., PATHAK S., KUAN L., JONES A., HAWRYLYCZ M., CARNINCI P., KASUKAWA T., KATAYAMA S., TOGA A.: Exploration and visualization of gene expression with neuroanatomy in the adult mouse brain. *BMC Bioinformatics 9*, 1 (2008), 153. 4

[LTW*11] LIN C. Y., TSAI K. L., WANG S. C., HSIEH C. H., CHANG H. M., CHIANG A. S.: The neuron navigator: Exploring the information pathway through the neural maze. In *IEEE Pacific Visualization Symposium 2011, PacificVis 2011 - Proceedings* (2011). doi:10.1109/PACIFICVIS.2011.5742370. 4

[MBB*16] MURUGESAN S., BOUCHARD K., BROWN J. A., HAMANN B., SEELEY W. W., TRUJILLO A., WEBER G. H.: Brain Modulyzer : Interactive Visual Analysis of Functional Brain Connectivity. *IEEE/ACM transactions on computational biology and bioinformatics* (2016), 1–14. 4

[mPlDI*16] MING POO M., LIN DU J., IP N. Y., XIONG Z. Q., XU B., TAN T.: China Brain Project: Basic Neuroscience, Brain Diseases, and Brain-Inspired Computing. doi:10.1016/j.neuron.2016.10.050. 1

[NSvA*13] NOWKE C., SCHMIDT M., VAN ALBADA S. J., EPPLER J. M., BAKKER R., DIESRNANN M., HENTSCHEL B., KUHLEN T.: VisNEST - Interactive analysis of neural activity data. *Biological Data Visualization (BioVis), 2013 IEEE Symposium on* (2013), 65–72. 3

[OHN*14] OH S. W., HARRIS J. A., NG L., WINSLOW B., CAIN N., MIHALAS S., WANG Q., LAU C., KUAN L., HENRY A. M., ET AL.: A mesoscale connectome of the mouse brain. *Nature 508*, 7495 (2014), 207–214. 2

[PLK*15] POLDRACK R. A., LAUMANN T. O., KOYEJO O., GREGORY B., HOVER A., CHEN M.-Y. E. A.: Long-term neural and physiological phenotyping of a single human. *Nature Communications 6* (2015), 8885. 3, 4

[SAM*05] SHERBONDY A., AKERS D., MACKENZIE R., DOUGHERTY R., WANDELL B.: Exploring connectivity of the brain's white matter with dynamic queries. In *IEEE Transactions on Visualization and Computer Graphics* (2005). doi:10.1109/TVCG.2005.59. 4

[SCHT09] SAALFELD S., CARDONA A., HARTENSTEIN V., TOMANČÁK P.: CATMAID: Collaborative annotation toolkit for massive amounts of image data. *Bioinformatics* (2009). doi:10.1093/bioinformatics/btp266. 4

[SE12] SCHMITT O., EIPERT P.: neuroVIISAS: Approaching Multiscale Simulation of the Rat Connectome. *Neuroinformatics 10*, 3 (2012), 243–267. 4

[Sor13] SORGER J.: neuroMap - Interactive Graph-Visualization of the Fruit Fly ' s Neural Circuit. In *Biological Data Visualization (BioVis), 2013 IEEE Symposium on* (2013). 4

[SSDV14] SCHOLTENS L. H., SCHMIDT R., DE REUS M. A., VAN DEN HEUVEL M. P.: Linking Macroscale Graph Analytical Organization to Microscale Neuroarchitectonics in the Macaque Connectome. *The Journal of Neuroscience* (2014). 3

[VESB*13] VAN ESSEN D. C., SMITH S. M., BARCH D. M., BEHRENS T. E., YACOUB E., UGURBIL K., CONSORTIUM W.-M. H., ET AL.: The wu-minn human connectome project: an overview. *Neuroimage 80* (2013), 62–79. 1

[XWH13] XIA M.-R., WANG J.-H., HE Y.: BrainNet Viewer: a network visualization tool for human brain connectomics. *PloS one 8*, 7 (2013). 4

[YW04] YOUNG L. J., WANG Z.: The neurobiology of pair bonding. *Nature neuroscience 7*, 10 (2004), 1048–1054. 2

© 2018 The Author(s)
Eurographics Proceedings © 2018 The Eurographics Association.

PAPER D

BrainTrawler: A visual analytics framework for iterative exploration of heterogeneous big brain data

Florian Ganglberger, Nicolas Swoboda, Lisa Frauenstein, Joanna Kaczanowska, Wulf Haubensak, and Katja Buehler. Braintrawler: A visual analytics framework for iterative exploration of heterogeneous big brain data. *Computers & Graphics*, 82:304 – 320, 2019

Computers & Graphics (2019)

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

BrainTrawler: A visual analytics framework for iterative exploration of heterogeneous big brain data

Florian Ganglberger[a,*], Nicolas Swoboda[a], Lisa Frauenstein[a], Joanna Kaczanowska[b], Wulf Haubensak[b], Katja Bühler[a]

[a] *VRVis Research Center, Donau-City Straße 11-13, 1220 Wien, Austria*
[b] *Research Institute of Molecular Pathology (IMP), Vienna Biocenter (VBC)*

ARTICLE INFO

Article history:
Received 15 January 2019
Received in final form 28 May 2019
Accepted 29 May 2019
Available online 7 June 2019
This is a post-peer-review, pre-copyedit version of an article published in Computers & Graphcis. The final authenticated version is available online at: https://doi.org/10.1016/j.cag.2019.05.032

Keywords: big data, networks, neuroscience, gene expression, brain parcellation, visual queries

ABSTRACT

In recent years, big brain-initiatives and consortia have created vast resources of publicly available brain data that can be used by neuroscientists for their own research experiments. This includes microscale connectivity data—brain-network graphs with billions of edges—and vast spatial gene expression resources—the representation of tens of thousands genes in brain space. Their joint analysis for higher order relations in structural or functional neuroanatomy would enable the genetic dissection of brain networks on a genome-wide scale. Current experimental workflows involve only time-consuming manual aggregation and extensive graph theoretical analysis of data from different sources, which rarely provide spatial context to operate continuously on different scales.

In this paper, we propose *BrainTrawler*, a task-driven, web-based framework that incorporates visual analytics methods to explore heterogeneous neurobiological data. It facilitates spatial indexing to query large-scale voxel-level connectivity data and gene expression collections in real-time. Relating data to the hierarchical structure of common anatomical atlases enables the retrieval on different anatomical levels. Together with intuitive network visualization, iterative visual queries, and quantitative information this allows the genetic dissection of multimodal networks on local/global scales in a spatial context.

We demonstrate the relevance of our approach for neuroscience by exploring social-behavior and memory/learning related functional neuroanatomy in mice.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

The quest for understanding the principle organization of the brain and its functional parcellation is constantly changing due to the increasing wealth of multimodal neurobiological data generated by brain initiatives, such as the Allen Institute [1], the Human Brain Project [2], the WU-Minn Human Connectome Project [3], and the China Brain Project [4].

Our work takes recent advances in circuit neuroscience into account (e.g. neuro- and behavioral genetics, optogenetics, imaging) that identified gene sets underlying a specific behavioral function [5]. However, there is a lack of tools to explore the mesoscale (i.e. scale of neuron populations) as well as the global structural and functional brain networks related to these gene sets in silico.

We meet this demand by proposing a task-driven, web-based framework called *BrainTrawler*. It allows the visualization, iterative exploration and the integration of spatial data on dif-

*Corresponding author: e-mail.: ganglberger@vrvis.at;

ferent scales or brain parcellations, including hierarchical and structural annotations of the brain, collections of 3D spatial gene expression data (i.e. which gene is where in the brain expressed) with tens of thousands images, and region/voxel-resolution networks with up to billions of connections. This enables workflows for interactively fusing volumetric with connectivity data, brain network exploration at different anatomical levels, genetic analysis of the brain and intuitive 3D graph visualization. We demonstrate the practical significance of this tool by presenting several use cases based on heterogeneous neurobiological data of the mouse brain from large-scale brain initiatives that allowed to reproduce several recent biological findings.

This paper represents an extension of our previous work [6]. There, we proposed a framework that visualizes volumetric, geometry, and graph data simultaneously in 3D rendering and 2D slice views, linked to views showing quantitative profiles at a hierarchical parcellation level. The user can interactively navigate the hierarchical levels, which provide spatial context by rendering their brain regions in 2D/3D views. We utilized a specialized data structure [7] that organizes and aggregates the voxel-wise connectivity data hierarchically. Via so called *Aggregation Queries* [7], which represent the aggregated incoming/outgoing connectivity of volumes of interest (*VOI*), the data is made accessible in real time and can be explored interactively on different anatomical scales. The connectivity is visualized as intensity volume in the 3D rendering as well as quantitative brain-region-wise profile. These are augmented with hierarchical information (i.e. intensity of sub-regions) to provide an overview of different hierarchical levels. This kind of interaction allows the researcher to relate integrated resources, for example incoming/outgoing connectivity at voxel-level and region level, directly to imaging data showing brain-wide gene expressions (spatial gene expression data). Higher order network connections can be targeted by repeatedly cascading over the query results, which can be selected directly in the brain-region-wise profile. This represents an iterative exploration of the networks. Region-level representation of voxel-level connectivity rendered as 3D and 2D network graphs can be used to reduce complexity and provides neuroanatomical context, so they can be presented in a way that is particularly suited for neuroscience publications, as advised by domain experts.

In this paper, we integrate the framework's components into a task-driven workflow scheme. It does not only allow the iterative exploration of big brain networks, it further enables the genetic dissection thereof (i.e. which genes play a role in specific parts of these networks). For this, we included spatial gene expression data of the mouse brain on a genome-wide level [8] (i.e. for all genes in the mouse genome) and used spatial indexing [9] to allow for real-time queries of genes expressed at a *VOI*. Query results can be analyzed in linked views that include visual analytic tools (parallel coordinates, heatmaps etc.) to get insight into genetic features of subnetworks, i.e. how genetically similar are different parts of a network or which genes are relevant for different connections.

Furthermore, we performed several design optimizations regarding the qualitative and quantitative visualization of networks. Quantitative information can be arranged in different ways to provide the user with either more detail or overview. In Ganglberger et al. [6] *Aggregation Queries* [7] could be represented as region-level graph in their 3D/2D spatial context. For this, the voxel-level connectivity of the *VOI* is mapped to a brain region-level. Since the result of an *Aggregation Query* represents the already aggregated connectivity of a *VOI*, it could not be further split into anatomical sub-regions thereof (i.e. the *VOI* represents the smallest anatomical level the outgoing/incoming connectivity can be mapped to). We overcame this by developing a *Split-Aggregation Query*, which divides a *VOI* to anatomical subregions and executes multiple *Aggregation Queries* in parallel. With similar performance to a single *Aggregation Query*, they allow a more detailed regional-level dissection of voxel-level connectivity in real-time.

Our previous work [6, 7] enabled us to utilize, for the first time, methods for interactive exploration of dense voxel-wise brain connectivity data of several dozen gigabytes of size in combination with vast genetic databases containing volumetric image data. Based on this, we propose visual analytics methods for interactive

- joint exploration and fusion of brain network data with spatial gene expression data continuously over different scales ranging from multi-resolution voxel-wise connectivity to connectivity defined in respect to different anatomical region hierarchy levels
- joint exploration of data from different brain network types (e.g. structural, functional, gene co-expression)
- identification of higher order connectivity in dense network graphs with billions of edges
- genetic analysis of brain networks based on a genome-wide gene expression database with tens of thousands images

These methods are based on a task-driven workflow that allows the sharing/storing its state for scientific data provenance.

2. Related Work

An abundance of visualization tools for neurobiological data exists, usually tailored to specific species such as the drosophila [9, 10, 11], mouse [12, 13], macaque monkey [14] or the human [15] or the type of data source such as 2D and 3D imaging data showing e.g. spatial gene expression [12, 16], anatomical structure derived from microscopy or medical imaging (traced neurons on electron microscopy data [17], diffusion tensor imaging fibers [18], anatomical regions and segmented neurons from light microscopy[9], different kinds of connectome data[19] and gene co-expression data ([20, 16]). While traditional visualization tools have to be installed locally [9, 21], there is a trend towards developing web-based services. They have the potential to integrate publicly available and individually administrable data repositories to facilitate integrated workflows, complying with the shift of brain connectomics into the "big data" era

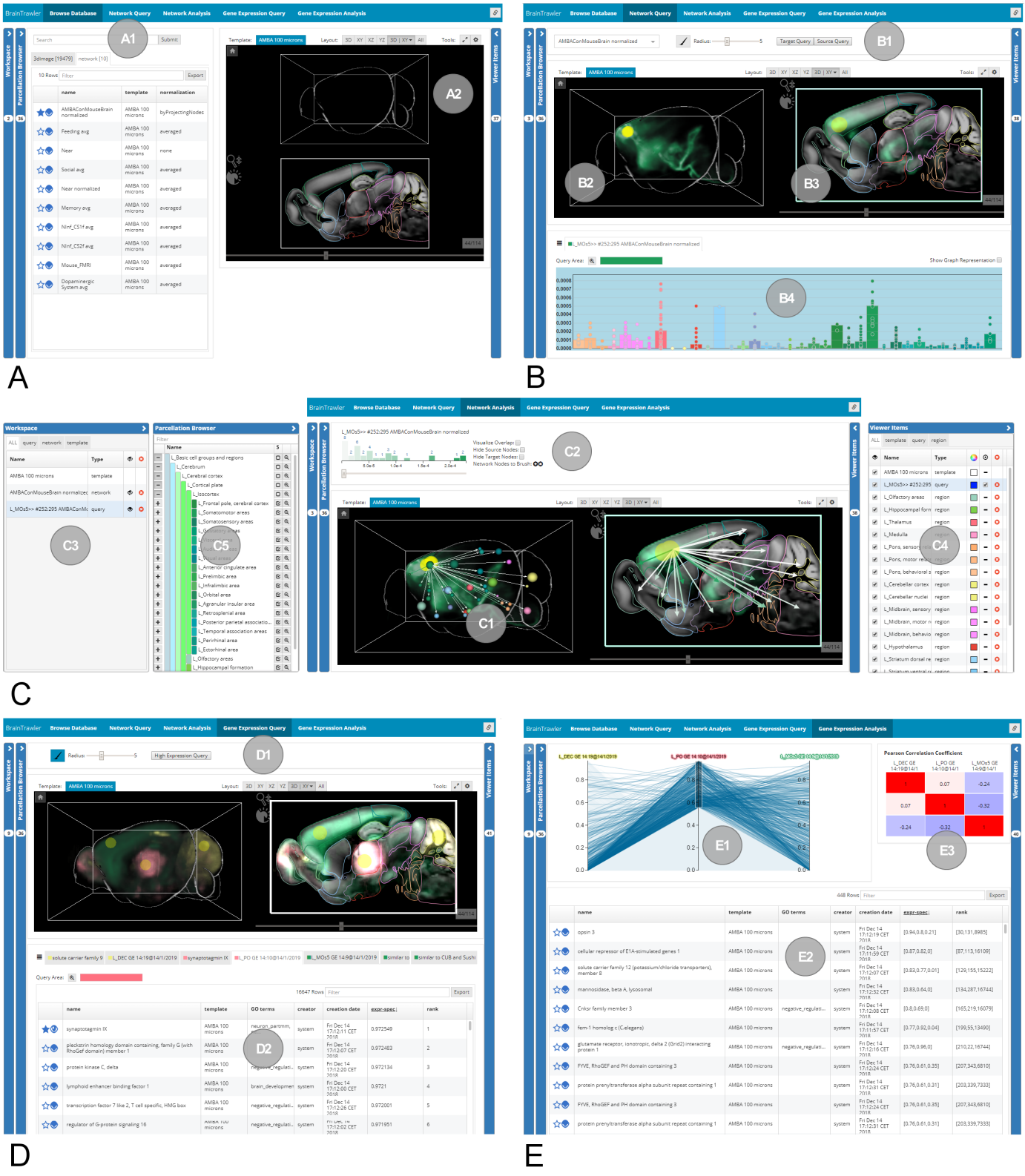


Fig. 1. Task-based web-component for iterative exploration of multiscale brain data. Detailed views of the individual tabs/subfigures can be seen in the supplement. A) Browse Database tab: The user can browse a collection of 3D gene expression and connectivity data via a text search (A1) and visualize the data directly in the 3D view (A2) or add it to the workspace (C3). B) Network Query tab: The user can execute target/source queries from the Query Toolbar (B1) on connectivity matrices that have been added to the workspace. Gene expression/Connectivity is rendered volumetrically in a 3D view (B2), in a 2D slice view (B3) and as region-wise quantitative representation (Expression/Connectivity Profile) (B4) C) Network Analysis tab: The user can explore networks or connectivities on a region level in a ball/stick representation (C1), controlled by the Network Analysis Toolbar showing histograms of the visualized graphs and threshold sliders for edge filtering (C2). The workspace manages which items (3D volumetric data, networks, query results) are shown in each tab (C3) while the Viewer Items List (C4) controls their appearance (e.g. color). The Parcellation Browser (C5) shows a hierarchical (anatomical) brain parcellation and is used to change the parcellation level of visualized graphs and profiles. D) Gene Expression Query tab: Via the Query Toolbar (D1), *Gene Expression Queries* can be executed. Resulting gene lists are shown in a table below (D2). E) Gene Expression Analysis tab: Gene lists can be compared in a parallel coordinate system (E1). Its selection is shown in a table (E2), its correlation in a heatmap (E3).

[22] and a complementation of experimental science by data driven science.

Joint exploration of neurobiological data on different scales: As various studies have already shown for the mammalian brain, combining information of the macroscale connectome with microscale neuronal architecture does provide a deeper understanding of the brain's organization [23]. Recently, this has also been confirmed for the living human brain, associating resting state functional connectivity with both gene expression [24] as well as cortical microstructure such as T1-based myelin content, as obtained from ultrahigh-resolution MRI [25]. This emphasizes the growing need for visualization tools which incorporate data of multiple scales. To visualize the relationship between multiscale data, mentioned studies primarily utilize enhanced heatmaps, as it can be seen in Scholtens et al. [23], where the matrix entries depict anatomical projections. Structural connectivity strength is encoded by color, projection distance by dot size. To provide spatial context, the matrix is presented side-by-side to a brain surface, color-coding anatomical areas. This approach provides a good overview on region-level. However its complexity increases on higher resolution, such as voxel-level.

Lin et al. [26] introduced Neuron Navigator, a tool that allows queries for connectivity of regions of interest in the Drosophila brain space. These spatial queries are achieved by accessing a 3D neuron image database and matching the region of interest with annotated locations of neuron terminals. Although these queries represent connectivity between neurons, they are not executed on actual graph/network data. Other tools allowing exploration of connectomic data are CATMAID [27] and ConnectomeExplorer [28], which allow tracing neurons on single EM Stacks, so that they work at a local level of a single network with fixed scale. Similar accounts for Sherbondy et al. [29], who query pre-computed pathways on diffusion tensor imaging data from volumes of interest. NeuroLines [30] offers a visualization as simplified skeleton graph, similar to a 2D subway map. Different levels of abstraction allow for multi-scale exploration, but is ultimately tailored to branched, tree like connections in electron microscopy data.

To integrate macroscopic data at a brain region level with microscale data from simulation of neural activity, Nowke et al. [31] have introduced the interactive analytics tool VisNEST. The tool provides views for visualizing connectivity between brain regions, within-region connectivity representation and time-varying activities across regions. Regions of interest can be selected in a 3D anatomical view and are represented as meshes, where color and opacity depict activity. Although VisNEST allows comparison of region-level connectivity, it is not tailored for high-resolution connectivity data nor does it relate to volumetric data.

A more generic framework, named neuroVIISAS [32] provides multiple ways of organizing, visualizing and analyzing multiscale brain network simulation data. NeuroVIISAS offers a collection of visualization techniques in 2D such as different representations of hierarchical connectivity matrices, circular connectograms, various layouts for planar graphs, as well as in 3D, where regions are rendered as spheres or true anatomical surface meshes, colored based on the color defined in the reference ontology. Connections are depicted as tubes with or without arrows. Moreover, individual source and target queries can be performed textually by defining filters on a table of all the available connections. The resulting selections can be viewed side-by-side in 2D atlas views. However, as this tool is designed for analysis of simulation data it has a different scope.

Recently, Huisman et al. [16] presented BrainScope, a web-based tool for interactive visual exploration of gene co-expressions and their anatomical context derived from the Allen Human Atlases offering a variety of linked views, where mouseover in one view highlights the according selection. Selections can be made by brushing regions and are saved in a workspace. Gene sets can be directly transferred to portals for gene enrichment analysis. They apply dimensionality reduction to explore the high dimensional data in two-dimensional space, represented in a scatter plot, providing spatial context by linking interactive views of atlas slices, and a hierarchical anatomy navigator, for selection and deselection of brain regions. BrainScope also allows the user to upload their own gene sets for exploration but is to date limited to the human template brains provided.

BrainExplorer [12] is an interactive visualization tool, provided by the Allen Brain Institute to explore the spatial gene expression and structural connectivity data provided by the Allen Mouse Brain Atlas. Brain Explorer provides volume rendering and blob-like visualization of gene expressions mapped to a standard brain. Voxels are colored according to their expression level. The tool enables the execution of pre-computed source/target connectivity queries at brain region level, which is also available via a web-interface on the Allen Institute's website [1]. Furthermore, BrainExplorer allows explicit gene queries, searching for specified genes of interest and anatomical queries where the search is based on specific anatomical regions of interest. For identified genes of interest, a correlation query can be executed, returning genes with similar expression profiles. This work comes closest to our solution. However, they provide (pre-computed) queries for incoming/outgoing connections only on pre-defined sites or anatomical regions.

Visualization of structural, functional and gene co-expression brain networks: Brain networks are commonly represented as node-link diagrams, connectivity matrices or connectograms, where only the first can be visualized in anatomical space both in 2D, with slice views of the anatomical planes and 3D. In volumetric space, edges have been rendered as straight lines in between nodes depicted as spheres across the brain [21, 19] or curved lines along the brain surface [33]. However, with a dense number of connections, this produces clutter and obscures the underlying anatomy. Mostly, brain networks are too complex to be able to show all details up to the highest level of resolution whilst seeing the entire global structure. Thus, hierarchies can define the level of partition. Applied to rendering in 3D space, the hierarchical level can be determined by the distance to the viewpoint in a level-of-detail-visualization [34], edge bundling [35]. Hence, brain network visualization tools [36] often present network structure and anatomical context side-by-side and achieve reference

through linked views and coloring network nodes according to their anatomical region. This provides the flexibility to layout the connectivity network according to network properties such as using a spring-embedded layout where well-connected groups of nodes are pulled together as seen in Poldrack et al. [24]. An interactive approach has been proposed in NeuroCave [37]. Instead of arranging nodes, they map the network to a topological space via dimensionality reduction of connectivity data. Here, it renderes a node-link diagram in anatomical (real world coordinates) and topological spaces side-by-side.

Different graph layouts, such as an anatomical layout, are also supported by NeuroMap [38], which renders potential neuronal connections in the fruit fly's brain as interactive circuit-style wiring diagrams. Spatial context is provided by introducing brain regions as spatial constraints to a 2D layout by providing a 2D abstraction of anatomical organizations and by linking NeuroMap with 3D visualizations, showing the neurons in their 3D anatomical context. Another anatomical layout approach has been used by Ji et al. [39], which uses a planar projection of the human skull visualize functional networks derived from EEG. Although these tools provide spatial/anatomical context, they are limited to a regional level, and do not scale to voxel-level resolution with billions of edges.

When comparing different kinds of connectivities, it is essential to visualize similarities/differences of networks [40]. A design study conducted by Alper et al. [40] showed that a matrix visualization in combination with glyphs results in a better comparability than superimposed node-link diagrams. Another way are so called small multiples, a series of graphs with similar scale to compare them easily, for example functional connectivity in a circular layout with similar anatomical regions/nodes [41]. For analyzing networks over time—with hundreds of time points—this approach becomes time-consuming and unreliable for it depends on memorization by the user. Therefore, Bach et al. [42] used a pilling metaphor to visually encode snapshots of a network (i.e. a network at a time point) into manageable parts ("piles" of similar small multiples) to identify temporal patterns in functional networks.

3. Background

3.1. Data

We have been collaborating with experts interested in neuronal circuits in the mouse brain, how they control emotional states and behavior, and how they are modulated by genes and psychoactive drugs [43]. The data they use provide a good starting point for understanding how nowadays data-driven research is done in neurocircuit science. It can be divided into three types:

Standard brain and its parcellation (building the reference system): A standard brain represents a common reference space, to which other data can be registered to. This standard space can be hierarchically parcelled to represent anatomical annotation (e.g. Allen Mouse Brain and Allen Human Brain region annotations). In general, the highest level is the whole brain which is divided hierarchically into sub-regions.

Volume data (aligned to the standard brain, i.e. everything is in the same reference space): 3D image data (e.g. brain template, spatial gene expression). The spatial gene expression is available on a genome-wide scale [8], therefore 3D image data of 19479 genes add up to to several dozen gigabytes in total.

Connectivity data (also aligned to the standard brain): Region-wise connections/relations (e.g. resting state functional connectivity) and voxel-resolution connections/relations (structural connectivity, gene co-expression) derived from imaging data (e.g. Allen Mouse Brain gene expression or connectivity data). Voxel-wise connectivity/relational data generally has a very dense connectivity with up to billions of connections. Hence, their matrices can take up hundreds of gigabytes.

3.2. User Stories

Based on previous projects, informal interviews, and discussions with our collaborating domain experts, we jointly identified the following major analytical workflows and user stories:

Relating spatial gene expression data to different kinds of connectivity: The entry point for many data analytics workflows are 'candidate regions'—brain regions that may be part of a specific brain circuit. Those are affected by genes that are either related to a certain behavior, or targeted by a psychoactive drug. Thus, the knowledge of where a gene affects the brain is a first step in relating it to a particular function. The effect is rather broadly defined. Well documented cases are primary gene expression sites [8] (sites where the gene creates products, such as proteins) and structural connectivity (regions to which primary gene expression sites project) [44]. Spatial gene expression and structural connectivity data for the mouse brain is provided e.g. by the Allen Mouse Brain Atlas (*AMBA*) at a voxel-level [8]. Here, the primary expression sites of a gene are made available as 8 bit intensity volumes, representing the density of cells expressing the gene (i.e. spatial gene expression data), related to a specific standard brain. To use those resources for hypotheses building, it is necessary to aggregate data manually by querying online databases and literature research. Therefore, an integrated workflow for data fusion is missing. Identifying brain regions that are functionally or structurally connected (either directly or transitively) to this expression site may also contribute to this function or be a second order effect thereof. Performing this tasks at a voxel-level is a particular challenge, since connectivity matrices can easily grow up to hundreds of gigabytes.

Exploring the data on different scales: Operating region-wise on the data (for example a region-wise network graph) depends on spatial hierarchy. Brain regions are organized by ontologies as trees with larger regions at the top, and ultimately resolves into voxel-resolution at the lowest level of the related standard brain. In order to keep an overview while comparing global networks with different modalities, larger regions are preferred. On the other hand, visualizing small subnetworks for circuit dissection requires smaller regions of the hierarchy or even voxel-level. A hierarchical organization of the network graph can therefore replace time-consuming pre-selection of relevant regions in current experimental workflows.

Comparing different types of connectivity is of essential importance for identifying neural circuits. For example, two

brain regions can have a high structural connectivity (a connection via neurons) but do not necessarily express the same genes (e.g. a so called ligand-receptor binding [45]). Therefore, their genetic connectivity for a specific gene set, represented as gene co-expression correlation (the correlation of the gene-expression between two regions) is low. Finding a correlation between regions in their fMRI activity may show their functional relation, but a difference in gene co-expression correlation could reveal that this activity relies on completely different molecular mechanisms. Directional structural connectivity can reveal the flow of information for (undirectional) fMRI connectivity [46]. The possibilities of brain network comparisons are versatile, but the given examples are in particular relevant for domain experts. Combining specific circuits for larger networks can be done again by manual data aggregation, such as whole brain fMRI studies, but require the expertise of a bioinformatitian.

Analyzing genetic diversity of brain regions and networks is necessary to identify their impact on neuronal function, and consequently on behavior. Finding genes that are specific for certain neural circuits is crucial in drug-discovery [43] and can therefore lead to novel treatments of diseases. Performing this search on a genome-wide level using e.g. spatial gene expression data provided by *AMBA* [8], represents an unbiased approach, which does not involve manual literature research for relevant genes prior to the analysis. Since the size of such spatial gene expression databases exceeds the memory limitations of current consumer level computers, data retrieval strategies for real-time querying are needed [9].

3.3. Task Analysis

Based on the user stories described in Section 3.2 we identified the following core tasks to be supported by our system. T1-T2 serve the first user story, T4 the second, T5 the third and T6-T7 the fourth. T3 and T8 represent general requirement for all user stories.

T1: 3D target/source query: *Which parts of the brain are connected to a certain volume of interest (e.g. a region with high gene expression)?*

T2: Higher order target/source query: *What is transitively connected to a certain volume of interest?*

T3: Anatomical context: *To which brain regions is my data related?*

T4: Explore network graphs: *How do my networks look like on different anatomical scales?*

T5: Compare network graphs: *How do my networks differ from each other?*

T6: Gene expression query: *Which genes are specific for a certain volume of interest (e.g. a part of a network)?*

T7: Explore genetic diversity of networks: *How similar are the specific genes for different parts of a network?*

T8: Reproducible/Shareable Workflows: *I want to share/discuss my findings, so what data did I use, and which analysis steps did I perform?*

4. Data Structures for real-time Queries of Big Brain Data

A prerequisite for the user acceptance of a visual analytics framework is its responsiveness during interaction. This task is in our case challenging, as we have to query large graphs, with billions of edges in real time. Furthermore, keeping the data, represented as connectivity matrices (weighted adjacency matrices) or collections of volume data in memory becomes infeasible with increasing size. This requires special solutions for fast disk access (T1, T2, T6). When operating on different anatomical scales (T3, T4), it is necessary to perform cumulative operations on the connectivity matrices (e.g calculate region-level connectivity from voxel-wise connectivity). In this case large parts of the connectivity matrix need to be loaded and processed.

4.1. Aggregation Queries for Big Brain Networks

We proposed a data structure in a previous publication [7] to allow these *Aggregation Queries* to happen in real-time, directly on the hard disk. It uses a specialized *Connectivity Storage* to efficiently manage and access large connectivity data (T1, T2). By exploiting the sparseness of the data and its spatial organization, it optimizes disk-reading speed via read-ahead paging and therefore achieves almost sequential reading-speed for local *Aggregation Queries* (details in Figure 2). To scale for larger brain areas, a cache mechanism loads pre-computed queries in a hierarchical way, similar to an image pyramid. Moreover, the *Connectivity Storage* provides a mapping to a standard reference space, which allows the retrieval of connectivity data regardless of its original resolution. The *Connectivity Storage* has been combined with a multi-model graph/document-database, which is tailored to hierarchical brain parcellations (i.e. hierarchical anatomical brain atlases) to store region-level connectivity for different hierarchy levels. It is also used to organize spatial data, such as volume data (e.g. spatial gene expression), binary masks, and meshes of region definitions.

4.2. Region-level Representation of Voxel-level Connectivity

The result of an *Aggregation Query* is a 3D image (*Connectivity Volume*), where the intensity represents the cumulated connectivity from a *VOI* (i.e. area in the brain) to the rest of the brain (i.e. every voxel). In our previous work [6], it was possible to visualize a region-level *Graph Representation* of a *Connectivity Volume*. This means a region-level representation of aggregated voxel-level connectivity (T4). Since this is the already cumulated connectivity of a *VOI*, the connectivity of anatomical subregions within the *VOI* could not be determined (i.e. one can not split the connectivity for subregions of the *VOI* anymore). It rather represented an abstract view of the query, showing the aggregated connectivity of a *VOI* mapped to a region-level in its 3D/2D spatial context (Figure 3A). However, exploring the connectivity of anatomical subregions provides a more precise view on the data, which would otherwise require additional queries by the user. We address this problem with *Split-Aggregation Queries*. Before executing a query, the *VOI* is split into anatomical brain-regions of the lowest level of a hierarchical parcellation. During the reading of connectivity data from the *Connectivity Storage*, we generate not only a

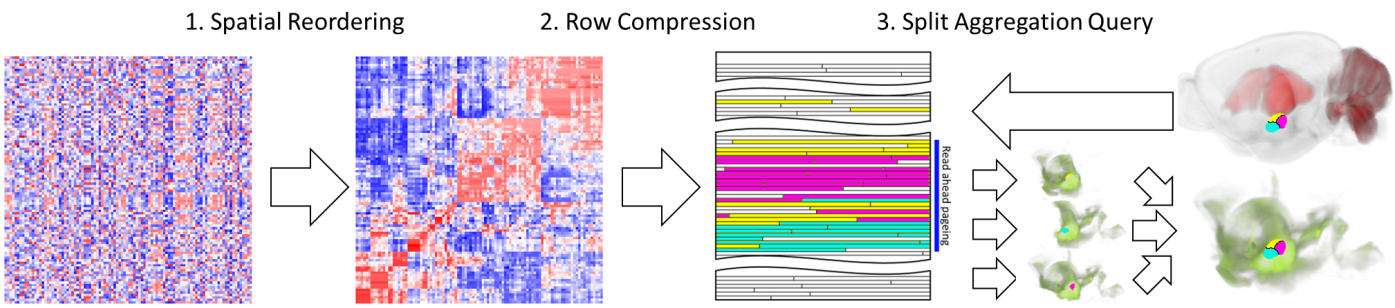


Fig. 2. Principle of the Connectivity Storage and real time Aggregation Queries: 1. Spatial reordering of a (voxel-level) connectivity matrix. Rows/columns that represent the outgoing/incoming connectivity of voxels are reordered, so that rows/columns that belong to spatially close voxel are close to each other. 2. Row Compression: The reordered connectivity matrix is stored with a row-wise compression as Connectivity Storage File on the hard disk. This saves disk space (and therefore reading time) by exploiting sparseness. 3. Split Aggregation Query: A *VOI* (yellow, purple and cyan indicates anatomical subregions of the *VOI*) is selected on 3D spatial data (e.g. spatial gene expression data, red). Outgoing connectivity is aggregated by reading the corresponding rows (yellow,red and cyan) from the Connectivity Storage File into three separate *Connectivity Volumes*. Since spatially close rows are also close in the file, they can profit from read-ahead paging (blue). Aggregated of the *VOI* can be generated from the three individual *Connectivity Volumes* (green) but is also available for the 3 subregions (e.g. for region-level graph representation).

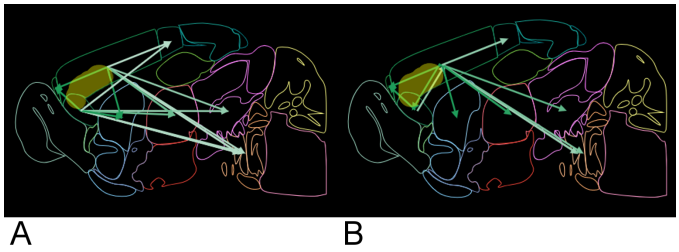


Fig. 3. Region-level *Graph Representation* of a *Connectivity Volume* with its *VOI* (yellow). A) The *Graph Representation* from our previous paper [6] maps the *Connectivity Volume* of the *VOI* to the two regions within the *VOI* similarly (i.e. so they receive the same outgoing connectivity). B) *Split-Aggregation Queries* generate multiple *Connectivity Volumes* for each region of the *VOI*, so the *Graph Representation* maps the *Connectivity Volumes* of the two regions individually (i.e. so they receive their actual outgoing connectivity.

single *Connectivity Volume*, but one for each region of the *VOI* simultaneously (indicated as yellow, purple and cyan in Figure 2). While the total connectivity of the *VOI* can be generated from the individual *Connectivity Volumes*, one can still compute the region-level connectivity for each brain region of the hierarchical parcellation, and therefore an exact region-level *Graph Representation* (Figure 3B).

4.3. Querying Collections of Spatial Gene Expression Data

Genetic dissection of brain regions/networks on a genome-wide scale requires querying large collections of 3D image data (T6, T7). For each gene of the *Allen Mouse Brain Atlas* [8], we generated a 8 bit spatial gene expression image with 1MB from the *AMBA API* [47], by scaling the gene expression to a range between 0 and 255. To account for regional specificity (i.e. genes can be expressed in the whole brain, which makes them not specific for *VOI*), we standard normalized (z-score) the gene expression of each gene. The normalized values are then scaled between 0 and 1 (i.e. 0 means a z-score $\leq$ 0 and 1 a z-score $\geq$ 2.5). To enable a time-efficient call for the genome-wide gene expression of a *VOI* as ranked list (i.e. genes ranked by their mean gene expression specificity in the *VOI*), we created a spatial index using space-filling curves similar to Bruckner et al. [9] allowing fast access of spatial data in arbitrary *VOI*s. While Bruckner stores spatial objects and their distances, we store gene expression per voxel across all datasets in the *Allen Mouse Brain Atlas*. Hence, a *Gene Expression Query* profits from read-ahead paging similar to *Aggregation Queries*.

4.4. Application Interface

A REST API is acting as the central access point for the data. It provides calls for querying the connectivity matrices and spatial gene expression, as well as importing them. Data import of connectivity matrices can be performed by providing the matrices in a row/column table format as csv. The data structure will then convert the data to a compressed file format for the *Connectivity Storage* and automatically aggregates region-level connectivity in the graph-database (see details in Ganglberger et al. [7]). It can then be used immediately afterwards for real-time target/source *Connectivity Queries*.

5. Workflow centric UI Design

We evaluated how to best adapt the existing workflow [6] for the analytical tasks (Section 3.3) by conducting an informal discussion about the usability with our domain experts. Extending the interface from our previous work [6] with new features was perceived to overload the interface. As a result we implemented a tab-based design (Figure 1) of the interface subdividing the eight tasks into logical sub-workflows (Figure 4). Although tabs can be switched arbitrary, their order reflects the principle experimental procedure of domain experts:

W1) Browse Database: This workflow (Figure 1A) represents the starting point for all other workflows. The user can browse the data collections and add it to a workspace (Figure 1C3). Transitions between these workflows, as well as their current state (visualized data, executed queries etc) are managed by this, which can be stored/shared as hyperlink to allow for data provenance and reproducibility (T8).

W2) Network Query: Here, the user can execute target/source queries (T1) or higher order target/source queries

(T2) to explore local brain network connectivity visualized in an anatomical context (T3) (Figure 1B). The user has 3D volume data that shows activity in certain areas, for example where a certain gene is expressed. She wants to know which areas are connected via structure or by function at voxel-level (T1). Hence, she can directly select the areas that are interesting for her based on a 3D brain-visualization and receives the results instantly.

It is possible to select different types of connectivities, which can be visualized similarly or by their overlap. Originating from the target/sources, the user wants to know what further connections exist, i.e., which brain regions (also on a voxel-level) are transitively connected to a certain area (T2). Therefore, she can perform target/source queries iteratively, aided by accurate *VOI* selection of regions with high connectivity.

3D volume data, such as spatial gene expression data or target/source connectivity query results need to be related to their anatomical context, so that the user has quantified information of what she is seeing (T3). Therefore, the volume data is summarized to show how much gene expression/connectivity individual brain regions exhibit. The hierarchical nature of the regions makes it necessary to visualize different hierarchical levels, the user can select a level based on her needs (Figure 1C5).

W3) Network Analysis: Local connectivity (as result of W2) or global whole-brain connectivity can be analyzed and compared as region-level networks on different anatomical scales (T3,T4,T5) (Figure 1C). The user wants to explore connectivity of experimental data, such as a fMRI network of a study. Hence, she executes target/source queries on the fMRI network (W2) and other modalities and renders a graph representation of the results (T4). Hierarchical brain parcellation can be navigated (T3), until the graph fits her needs. Multiple connectivities are represented as a graph for purposes of comparison, i.e. to see the difference from and the overlap with her experiment (T5).

W4) Gene Expression Query: In this workflow (Figure 1D), gene expression queries (T6) can be performed. Here, user's research interest is a specific brain region or network node (i.e. a *VOI* that represents parts of a network from W2 or W3). She wants to know which genes have a spatial expression pattern specific for this region/node, which can be selected in a 3D brain-visualization. A list of genes, ranked by their gene expression specificity (i.e. how specific is gene expression of a gene for a *VOI*) is retrieved in real-time and shown as table with additional information (such as the genes functional association).

W5) Gene Expression Analysis: Here (Figure 1E), the user can explore the gene expression for different brain regions (e.g source/target regions of a network), i.e. she wants to know how similar these brain regions (T3) are genetically (T7). For this purpose, she can perform multiple gene expression queries in W4. The resulting gene lists can be compared and filtered interactively in a parallel coordinate system, based on the gene expression specificity and functional associations.

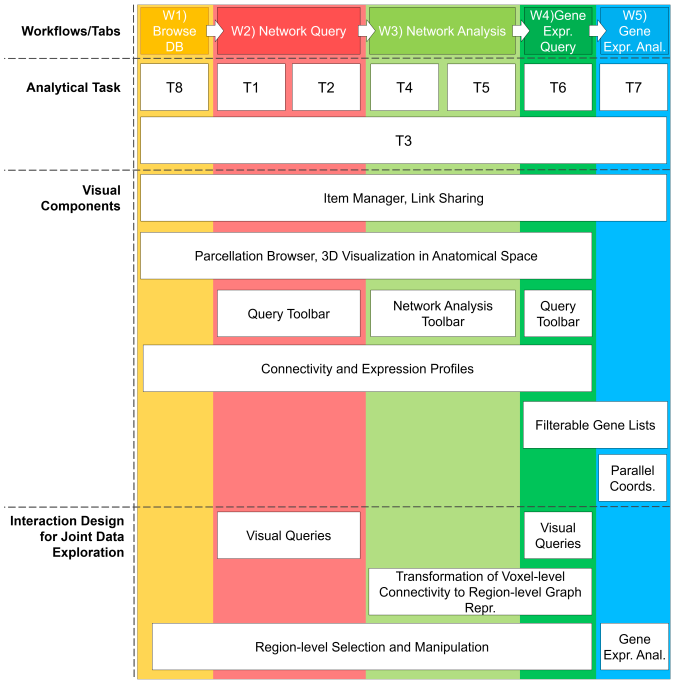


Fig. 4. Association of the workflows/tabs (Section 5), analytical tasks (Section 3.3), visual components (Section 6), and interactions for joint data exploration (Section 7). The components and interactions are arranged by their order of appearance in the main text.

6. Visual Components

This section describes the visual components that are used to implement the workflows (Figure 4) for the 8 analytical tasks described in Section 3.3.

6.1. Item Manager

The *Item Manager* is used to handle the association, visibility, and appearance of items between all workflows/tabs and is in particular important to provide ordered and reproducible workflows (T8). We split the *Item Manager* into two separate lists to emphasize the distinction between what is relevant for a tab (*Workspace List*) and the visual appearance within a workflows/tabs.

The *Workspace List* (Figure 1C3) manages which items are visible in which workflow/tab. Since not all items are relevant for all workflows, this supports the user to focus on her specific task. For example, spatial gene expression data of certain genes might be relevant as starting point for *Connectivity Queries* (Figure 1B, W2) but irrelevant for Network Analysis (Figure 1C, W3).

The *Viewer Item List* (Figure 1C4) controls the visibility and appearance of spatial volumetric data, brain regions and connectivity. For volumetric data the brightness, contrast, and transparency can be controlled to improve visibility.

6.2. Link Sharing

The current state of the application (visualized items, workspace, selection, executed queries, etc.) can be exported as hyperlink via a button on the top right of the application.

The user can store and share her current analysis for data provenance/reproducibility in a neuroscientific environment (T8).

6.3. Parcellation Browser

The *Parcellation Browser* (Figure 1C5) shows a hierarchically organized brain parcellation in a tree view, where every brain region has a name and a color code. This color-coding supports the user in relating brain regions to all other visual components. Its main purpose is to define the brain regions that are visualized (T3) in the *Expression and Connectivity Profiles* as well as the query result's *Graph Representation* (the *selection state*). The uppermost regions are the two hemispheres that can be expanded further down the hierarchy. Each major brain region has a color code that is varied in shade depending on its hierarchical level, providing the user with an indirect reference to the larger anatomical context of a subnode and to the spatial position in the brain. This effect is further enhanced by using established color codes for brain regions together with brain ontologies, which, for example, available in the *Allen Brain Atlas*.

6.4. 3D Visualization in Anatomical Space

We use classical volume rendering (Figure 1B2) and multiplanar reformation (Figure 1B3) to visualize volumetric data in its original spatial, anatomical environment (W1-W4). Graphs are rendered as ball-and-stick models, since they are common in the neuroscience community. We further render anatomical context in the planar view, visualized by region contours, and mesh geometry in 3D (if enabled via *Viewer Items List*) in order to provide anatomical context. 3D mesh rendering is turned off by default, since it would obstruct 3D volume rendering.

6.4.1. Volume Visualization

Volumetric data—gene expression data or voxel-level structural connectivity (T1, T2)—is rendered in 3D (Figure 1B2), and in three planar slice views (Figure 1B3) showing outlines of anatomical regions (T3). A maximum of four volumes can be blended simultaneously in 3D using single color-based transfer-functions (limited by the four RGBA channels, since the volume is loaded as texture). In the slice views, blending is also used for multiple images. Therefore, the rendering of intensity-overlap converges to white.

6.4.2. Network Visualization

A result of a target/source query can be also visualized as a 3D ball-and-stick-model within its anatomical reference space (T4, T5). Network nodes are rendered as spheres, where their diameter is determined by the size of the corresponding anatomical region. Color reflects their anatomical representations in the originating parcellation. Rendering them as meshes instead, would cause blocking the view by larger brain regions. We chose this ball-and-stick model in particular because they are common in neuroscientific publications (e.g. [48, 49, 50]). Hence, they are intuitively known in the neuroscientific community and can therefore used to share/discuss findings with collaborators that are not familiar with *BrainTrawler* (T8). For directed networks, edges are rendered as arrows, undirected network edges are represented by lines. Rendering arrows that are indicating the direction of connections were specifically requested by our expert users as they are an intuitive way of differentiating incoming from outgoing edges in publications. For weighted networks, connectivity strength is color-coded, ranging from white to the color selected in the *Viewer Item List* so different networks can be visually distinguished. Edges and nodes are user-adjustable via color-transfer-functions. Moreover, thresholding edge weights enables dynamical network refinement and an exploration of connections of interest. Node labels further support orientation.

In addition to the network model in the 3D view, network regions are shown as colored outlines of their anatomical equivalent in the corresponding 2D slice views; arrows (for directed) or lines (undirected networks) are rendered between the region centers in every slice. Although this shows only a part of the network (i.e. in the specific slice), it has the advantage that one can correctly render the regions by anatomy (instead of using a sphere), without facing the problem caused by possible occlusions in 3D.

The comparison and joint exploration of multiple graphs, which relate to the same anatomical parcellation, has been realized by overlaying several graphs, i.e. simply rendering multiple edges between nodes (Figure 6C1) for it is intuitively understandable. Since showing only two graphs even renders up to four arrows between nodes, we implemented an overlap visualization to emphasize on connections that are strong in multiple networks. Here, only a maximum of two edges per node needs to be rendered (i.e. two for directed, one for undirected). Their weights are defined by the multiplication of all edges (at region-level) between those nodes. Unlike an overlap that is based on the presence of connections (i.e. it renders binary edge weights if they are above certain thresholds in different networks), it provides edge weights on a continuous scale to visualize contrast between weak/strong connections. If an overlap is computed, the histograms depict a formula of the multiplication of individual graphs (Figure 6C2), so that the calculation can be comprehended by the user. Cascading queries are treated as a single network, depicted by brackets in the formula (see Section 7.2), since they represent connectivity of different orders.

6.5. Query Toolbar

The *Query Toolbar* (W2, W4) (Figure 1B1 and D1) can be used to execute target/source *Connectivity Queries* on connectivity matrices (T1, T2) as well as *Gene Expression Queries* (T6). Furthermore, it can start brushes, which is a spherical selection tool in the 2D slice view.

6.6. Network Analysis Toolbar

Since the number of edges increases quadratically with the number of nodes, the user can apply thresholding on edges to hide weaker connections (T4,T5). This is done with a slider to set a user defined threshold, in combination with a histogram that shows the edge weight distribution in the *Network Analysis Toolbar* (Figure 1C2, W3). The colors of the histogram bars map directly to the visualized edges. Hence, color directly relates to connectivity strength. The *Network Analysis Toolbar* further allows to hide source or target nodes of a network if the user wants to focus on these individual parts (T4).

6.7. Connectivity and Expression Profiles

Connectivity Profiles as well as *Expression Profiles* (Figure 1B3) summarize volumetric voxel-level data in order to make them quantitatively comprehensible for the user (T3). They visualize the mean target/source connectivity or gene expression for brain-regions in bar graphs. These brain regions are defined by the selection made in the *Parcellation Browser*. The mean connectivity/expression of all subregions are rendered as dots. This allows the user both to identify regions that have highly connected/expressed subregions, and to further refine the anatomical hierarchy to focus on relevant subregions. For anatomical context, the colors of the bars and dots are corresponding to the *Parcellation Browser*. Names are made visible via a tooltip on demand, since a direct rendering with the bars would cause clutter. One or multiple regions can be picked for a high-intensity *VOI* selection, which enables the user to threshold for high expression/connectivity voxels in a separate dialog (see Figure 5). This allows higher order target/source queries in an iterative workflow (T2), as further described in Section 7.1. For *Connectivity Profile*s, the percentual composition of the query region is shown in a bar above the profile to indicate the queries anatomical context. It uses the same parcellation/color code as the profile, and reveals its region names via mouse-over. Furthermore the query region can be added again to the 2D/3D viewer, so that the user can reproduce the query with the same or other connectivity matrices.

Profiles can be visualized in an overview next to the 2D/3D view to show multiple profiles below each other (Figure 7A). Therefore, regions are aligned vertically which enables comparability between multiple profiles (T5).

6.8. Filterable Gene Lists

Gene lists (Figure 1A1, D2 and E2, W4 and W5) show tabular information about spatial gene expression data by their name, *GO terms* (Gene Ontology, i.e. functional associations [51]), creation data, and—if they are result of a *Gene Expression Query*—their mean gene expression specificity. The list can be filtered for genes that have specific functional associations (T6, T7) textually. This is necessary since result lists may contain irrelevant genes that are not in the research focus of the user.

6.9. Parallel Coordinates System

A parallel coordinates system allows the filtering of multiple gene lists from different brain regions (W5) by their gene expression specificity (Figure 1E1), while their correlation is visualized in a heatmap (Figure 1E2) (T7). We chose a parallel coordinates system so one can interactively select genes that have a high specificity in one region, but low specificity in others (this is relevant for ligand-receptor systems). Furthermore, one could adapt this system for future projects, for example highlighting specific subsets of genes (e.g. related to a brain function or behavior) in a different color, so one could directly see the distribution of these genes in different brain regions.

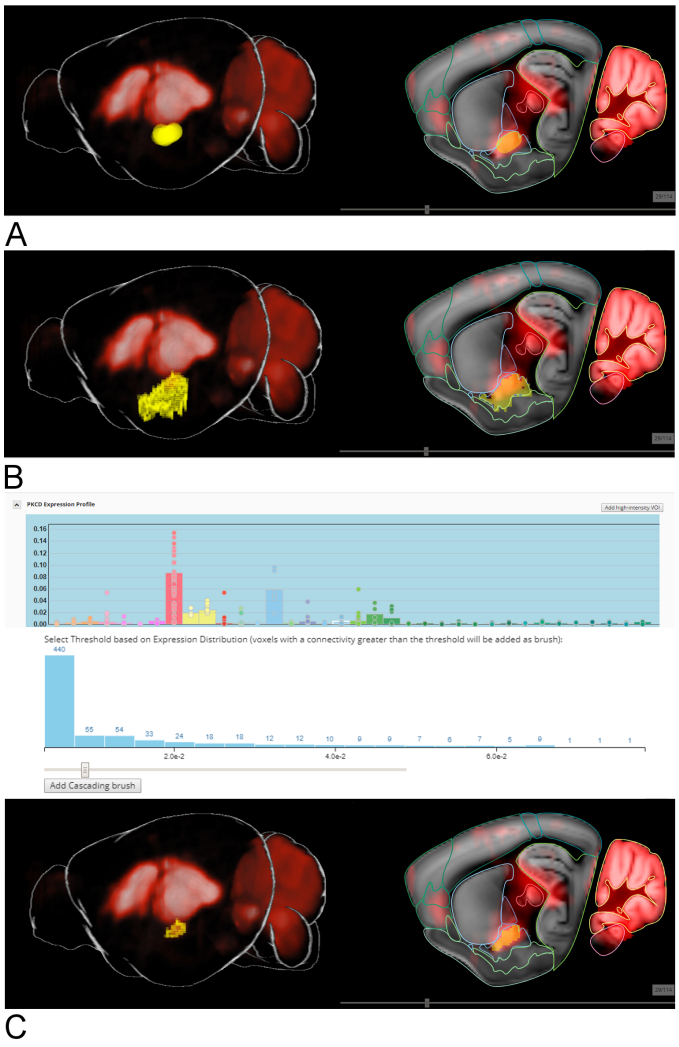


Fig. 5. Selection of parts of a gene expression pattern (red). A) Brush-selection of spherical region (yellow) in 2D slice views and rendering in 3D. B) Region-selection initiated by the Parcellation Browser, voxel-level visualization of a region (yellow). C) High-intensity-selection of a brain-region started from an *Expression Profile*. A histogram of voxel-level connectivity within the selected region lets the user decide which intensity, and therefore which voxels, will be selected. The result is rendered in 2D and 3D (yellow).

7. Interaction Design for Joint Data Exploration

7.1. Visual Queries

We use the Spatial Visual Query [9] paradigm to realize connectivity and gene expression queries. A visual query allows API requests to the data structure that are based on selections of a volume of interest directly in volume rendering. The response can then be shown in the 2D/3D renderer/*Connectivity Profile* (for target/source *Connectivity Queries*, T1, T2) or as gene list (for *Gene Expression Queries*, T6). This subsection describes the interaction with volume data, and the query types we have implemented.

7.1.1. Selecting a Volume of Interest:

In the 3D visualization (Section 6.4), the selection of areas can be performed in four different ways: Brush-selection,

region-selection, network-nodes-selection or high-intensity-selection. Brush-selection is performed from the *Query Toolbar*. It lets the user draw spherical areas encoded in transparent yellow in the 2D slice views, which are also directly rendered in 3D. For example, Figure 5A shows a gene-expression volume, where the spherical area is drawn on voxels with high gene-expression. Executing queries on this area is therefore acting as link between volume and connectivity data. The region-selection (Figure 5B)—added by the magnifying-glass-plus-button in the *Parcellation Browser* next to a region—selects all voxels within this region rendered as transparent yellow cubes. This allows the user to explore a network without volume data (e.g. if now relevant volume data is available). It is also possible to select the nodes of networks (i.e. the regions that represent the nodes of a network). This is similar to selecting the individual regions of a network manually, but it can be done in a single step W3. High-intensity queries are started from the *Expression Profiles*. They allow the user to select voxels with high intensity within user-selected brain regions. The user selects brain regions of interest in the *Expression Profiles* (Figure 5 C) with high gene expression. By clicking the "add high-intensity VOI" button, a dialog appears which shows a voxel-level histogram of the intensity values within this region. Here the user defines high-intensity voxels by setting a threshold. In the 3D/2D views, the voxels are instantly selected and visualized, similarly to the region-selection.

7.1.2. Target/source Connectivity Queries

Target/source Connectivity Queries (W2) can be used to link connectivity data with volume data (T1). In the *Query Toolbar* (Figure 1B1), connectivity matrices for querying can be selected, and brush-drawings with a certain radius can be initiated (although one can also use the other brush-types instead). The selected area (Figure 5A) then acts as an input for a target or source query on the API. The API retrieves the connectivity to all voxels that are either targets or sources from the selection, and then returns a *Connectivity Volume* (connectivity from/to VOI on voxel-level) as compressed JPEG to the web-component. It will be rendered instantly in 3D, 2D slice view, and as *Connectivity Profile* (T3). This represents the cumulated connectivity to (target) or from (source) the selected area. The *Connectivity Volume* gets automatically assigned with the name and color of the largest anatomical brain region in the *VOI*. Hence, it is easier for the user to associate the item with the query, than using a random name/id. *Connectivity Profiles* are shown directly below the 2D/3D view (Figure 5B4). Different tabs allow the user to switch between profiles of visible queries/gene expressions.

7.1.3. Higher order target/source Queries:

On the *Connectivity Profile* of the *Connectivity Volume*, the user can select regions of interest with high connectivity to start a high-intensity-selection (Figure 5C). Similarly to the *Expression Profile*, the user can choose a threshold in a voxel-level connectivity histogram, which adds all voxel within the region, and a connectivity above the threshold, as *VOI*. After performing a target/source query on the selection, the resulting second-order connectivity is visualized in 2D and 3D as another volume, while the associated *Connectivity Profile* is rendered again in a tab below the 2D/3D view. The profiles can be rendered in an overview (Figure 7A) were higher order *Connectivity Profiles* will be shown below their originating *Connectivity Profile* (T2). The name is indented and prefixed with "second-order" in red font, so that it is clear to the user that this query belongs to an iterative cascade. The cascading nature of these queries is further emphasized in the region-level graph representation, described in the following section.

7.1.4. Gene Expression Queries

Gene Expression Queries can be used for genetic dissection of brain regions or networks (T6) in W4. Similarly to target/source queries, they can be executed from *VOIs* selected by brush, regions, network-nodes, or high-intensity. The API retrieves the the mean gene expression specificity of the *VOI* for every gene in the spatial index, and renders it instantly as *Filterable Gene List* below the 2D/3D view (Figure 1D2) in a tab similar to *Connectivity Profiles*. The name and color of the gene list item (i.e. in the *Workspace* and *Viewer Item* list) are automatically assigned analogous to target/source connectivity queries (i.e. name/color of the largest anatomical region in the *VOI*). This is especially helpful for *Gene Expression Analysis* (W5), since it provides the user with anatomical reference without 2D/3D visualization (Figure 1E)

7.2. Transformation of Voxel-level Connectivity to Region-level Graph Representation

Changing from voxel-level *Connectivity Volumes* to region-level graph representation can be controlled either directly from a *Connectivity Profile* via "Show Graph Representation" button, or via the *Viewer Item List* (T3,T4,T5) in W2 or W3. It allows iterative queries for higher order target/sources to be visualized in a cascading way, since the individual graphs are connected by arrows (Figure 7B). Edges of the graph are drawn from all regions in the query area, to all regions in the *Connectivity Profile* (i.e. the selection state), with their mean connectivity as weights. These edges do not need to be calculated, since they have already been aggregated for the *Connectivity Profiles*.

7.3. Region-level Selection and Manipulation

The central control element for the *Connectivity Profile* and *Graph Representation* is the *Parcellation Browser*. Navigating in the tree directly influences the selection state, thereby determining which bars (profile) or nodes (graph) are shown (T3). Opening a subtree of a region will remove the corresponding bars/nodes from the graph, and adds all its children. Closing a region will add it, while removing its children. Since the user is not limited by a rigid parcellation, she can focus on regions of the brain that are relevant to her (e.g. only sub-cortical areas).

7.4. Gene Expression Analysis

A gene list as results of a *Gene Expression Query* reveals which genes are specific for a certain *VOI* (T6). However, multiple queries are needed to dissect circuits or a network (i.e. they involve different brain regions and therefore multiple *VOI*s)

(T7). Since every gene list consists of several thousand genes, an efficient comparison and combined filtering of these lists is needed. This can be done with a parallel coordinate system (Figure 1E1) in W5, where every coordinate shows the gene expression specificity for each *Gene Expression Query*. Query name and color provide the user with spatial reference (T3). Brushed genes in the parallel coordinate system are automatically listed below in a table (Figure 1E2), while their correlation is shown in a heatmap (Figure 1E3). To further narrow down the search, the list can be textually filtered, for example to look only for genes with specific functional associations.

8. Case Studies

We conducted three case studies in collaboration with our domain experts who also co-authored this paper. The studies represent use cases that would appear in a neuroscientific environment in this or a similar form. The first case study compares different kinds of networks (structural-connectivity vs gene co-expression correlation) related to social-bonding behavior (T1, T3, T4, T5). The second examines neurocircuits related to memory and learning, which we explore iteratively (T1, T2, T3). The last case study verifies the validity of the gene expression index by genetically dissecting a well known connection in the limbic system (T3, T6, T7).

8.1. Comparison of multiple networks

In the first case study our domain experts would like to examine brain networks related to social-bonding behavior. In particular, they were interested in oxytocin and vasopressin (neuropeptides known to be related to social behavior in mammals) release effect at the network level [52]. For this, they wanted both to examine primary expression sites of both genes, and explore their target sites (i.e. outgoing connectivity of expression sites) on different networks. Following data has been incorporated for this case study: spatial gene expression data (67x41x58 volume on a 200-micron resolution) of oxytocin (OXT) and vasopressin (AVP), a spatial gene co-expression correlation network (also 200-micron resolution, matrix file size is $\sim$ 12 GB) of social-bonding related genes consisting of (mouse gene ENTREZ ID in brackets) Oxt (18429), Oxtr (18430), Avp (11998), V1b receptor (26361), D1R (13488), D2R (13489), Slc6a3 (13162) and Crh (12918); and a 100-micron voxel-level (132x80x114) structural-connectivity from the *AMBA* (file size is $\sim$ 90 GB) [7]. Although these networks have billions of edges, our data structure [7] allows real-time retrieval of aggregated connectivity (i.e. cumulated incoming/outgoing connectivity) of a *VOI* on a 132x80x114 standard brain space. The user adds this data to the workspace in the *Browse Database* tab (W1) before the case study.

Identify a volume of interest of spatial data: The entry point for experts is the spatial gene expression data of OXT (Figure 6A cyan) and AVP (Figure 6A purple). The user notices an overlap in the hypothalamus (indicated by dark red contours, in the 2D slice view) (T3). She selects the overlapping area with a spherical brush initiated by the *Query Toolbar* (Figure 6B yellow spheres, highlighted by red arrow).

Find connected areas: In the *Network Query* tab (W2), the user selects the structural connectivity matrix in the *Query Toolbar*, a click on the *Target Query* button executes the selected *VOI* on the data structure (T1). The accumulated connectivity is instantly displayed as *Connectivity Volume* in 3D and 2D (Figure 6C green) and as *Connectivity Profile* (Figure 6C bottom). The profile shows that the strongest connectivity is to the Hypothalamus (dark red) itself (strongest connectivity to itself is common for structural and gene co-expression correlation). Other top connections are the Striatum (light blue), Pallidum (dark blue), Midbrain (pink), Thalamus (light red) and some cortical areas (green) (T3). This process is repeated for the gene co-expression correlation (Figure 6D blue), where its profile depicts hypothalamus as strongest again, but among similar connections also strong connections to Cerbellum (yellow) and Hindbrain (orange).

Compare networks: To compare the connectivities, the user visualizes them as *Graph Representation* (T4,T5) in the *Network Analysis* tab (W3). Figure 6E shows the two graphs in 2D and 3D, easily recognizable by their colors. For the user, the graph looks too cluttered, so she selects a threshold to filter weak connections in the *Network Analysis Toolbar*. Among others, the user sees that both networks show projections to the Septal Complex (LSX), the Anterior Cingulate (ACA) and the Prelimbic/Infralimbic Area (PL and ILA), which are well known nodes involved in social behavior [52]. To further highlight this, the user also visualizes the overlap of both graphs (i.e. multiplication of both connectivities). The resulting reduced amount of edges makes it easier for the user to identify regions which are connected via both connectivities (T5). These types of network visualization were familiar to our domain experts, who confirmed their suitability for neuroscientific visualization or publications.

This case study illustrates the joint exploration of gene expression data and different kinds of connectivities on voxel- and region-level. This approach allows real-time visual analytic workflows which are fast and efficient, compared to time-consuming manual data aggregation by querying different online databases, literature research and scripting. A video of the case study is available as *Supplementary Video 1*.

8.2. Higher-order connectivity

We designed this case study with our domain experts to showcase iterative higher-order *Connectivity Queries* with a well-known relationship between the brain-derived neurotropic factor (BDNF) and hippocampal synaptic plasticity, respectively circuits related to learning and memory [53]. One of these circuits is Dentate Gyrus (DG) $\rightarrow$ CA2/CA3 $\rightarrow$ CA1 $\rightarrow$ Entorhinal cortex/Subiculum [53]. For this purpose, we explored a primary expression site of BDNF in DG, and traversed their first-, second- and third-order targets iteratively. This case-study required the spatial gene expression data at a 200-micron resolution for BDNF (mouse gene ENTREZ ID: 12064) and structural connectivity at a 100-micron resolution from *AMBA* in the data structure [7], which is added to the workspace in the *Browse Database* tab (W1).

Identify a volume of interest of spatial data: The experts

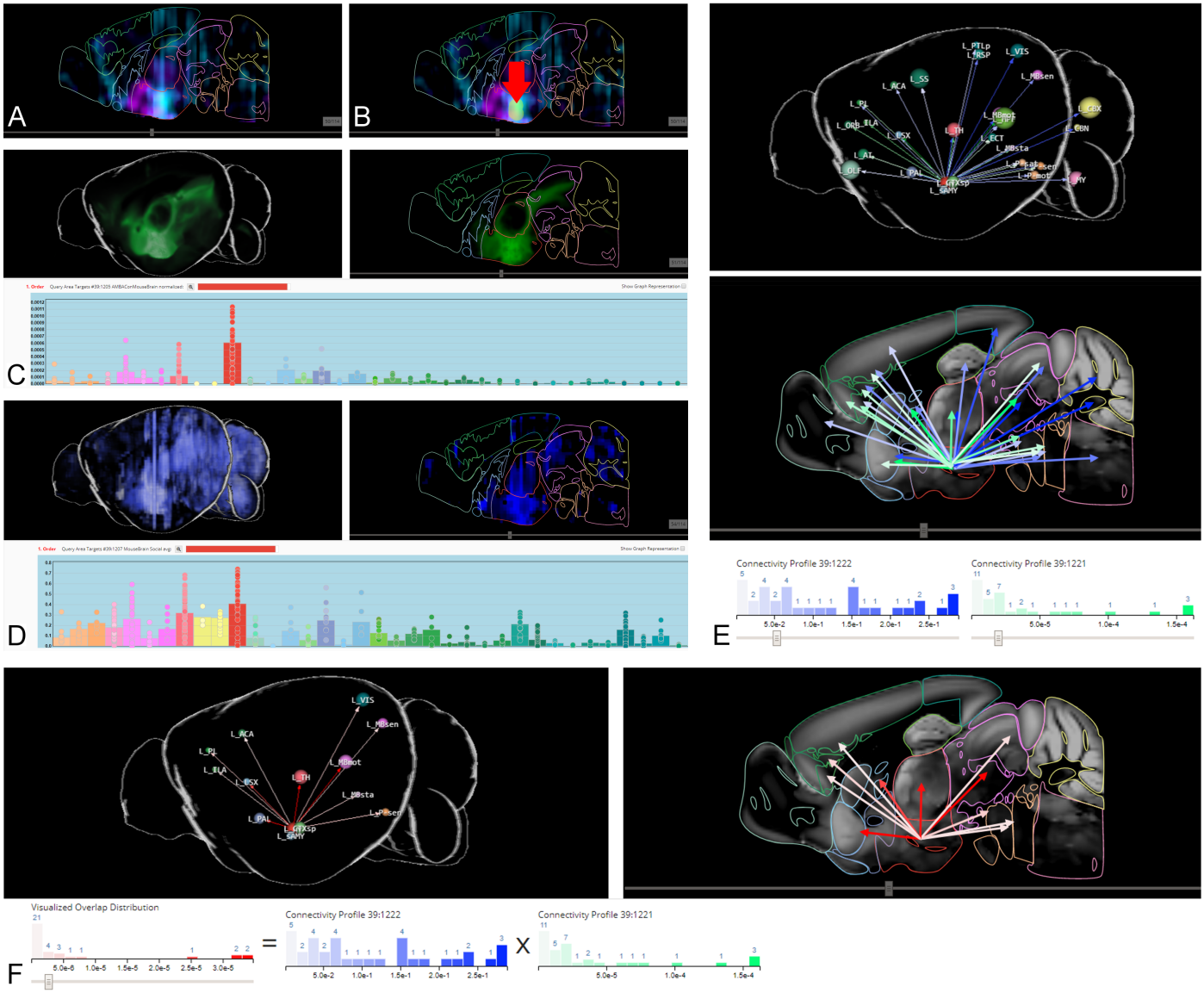
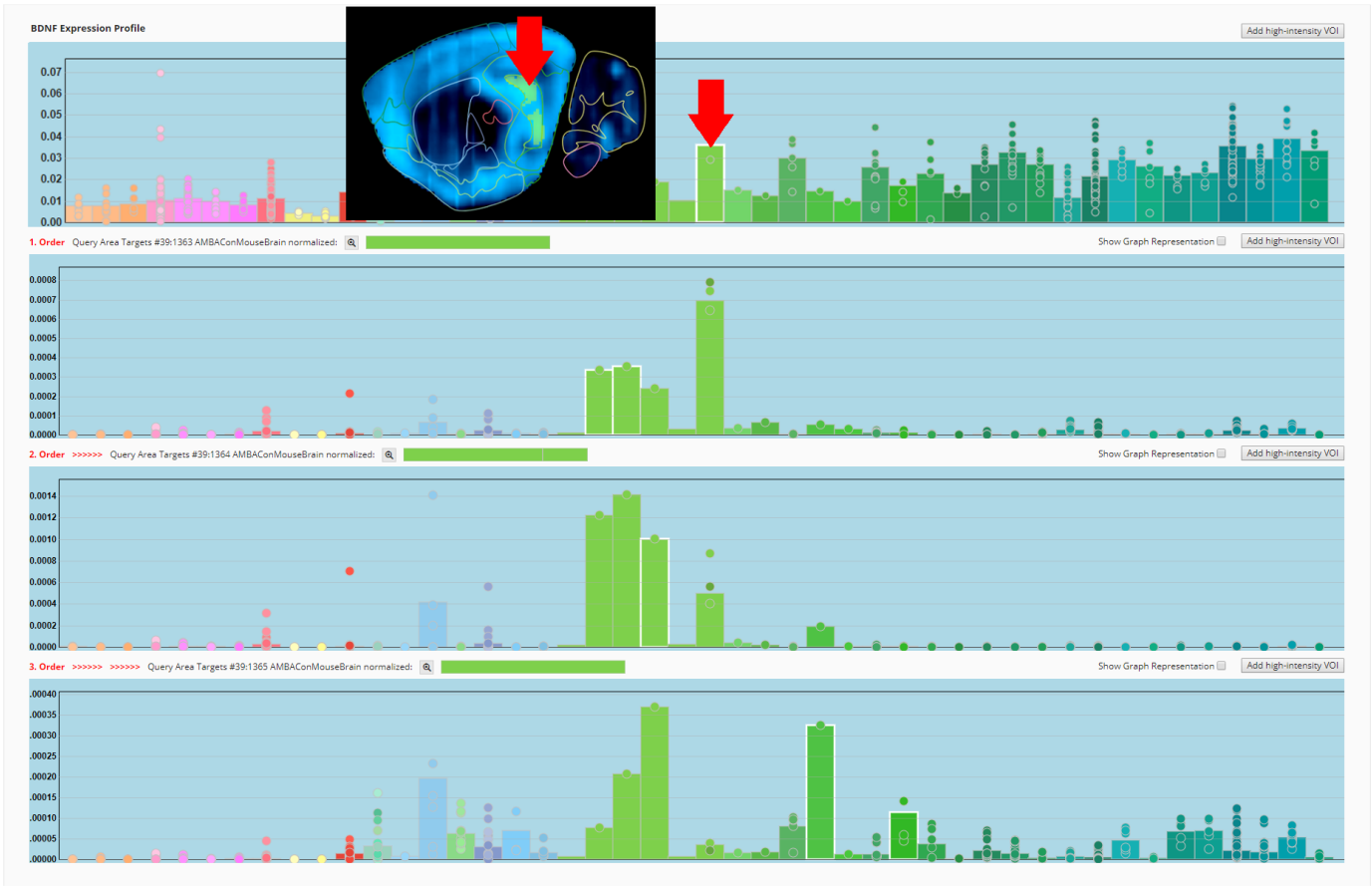


Fig. 6. Case Study 1: A) Gene expression of OXT (cyan) and AVP (purple). B) Selected VOI with brush-selection (yellow, indicated by red-arrow). C) Structural connectivity of the VOI in 3D and 2D (green) and its Connectivity Profile. D) Gene co-expression correlation of the VOI in 3D and 2D (blue) and its Connectivity Profile. E) Graph representation of structural connectivity (white to green) and gene co-expression correlation (white to blue). F) Multiplication (overlap) of the connectivities (red).

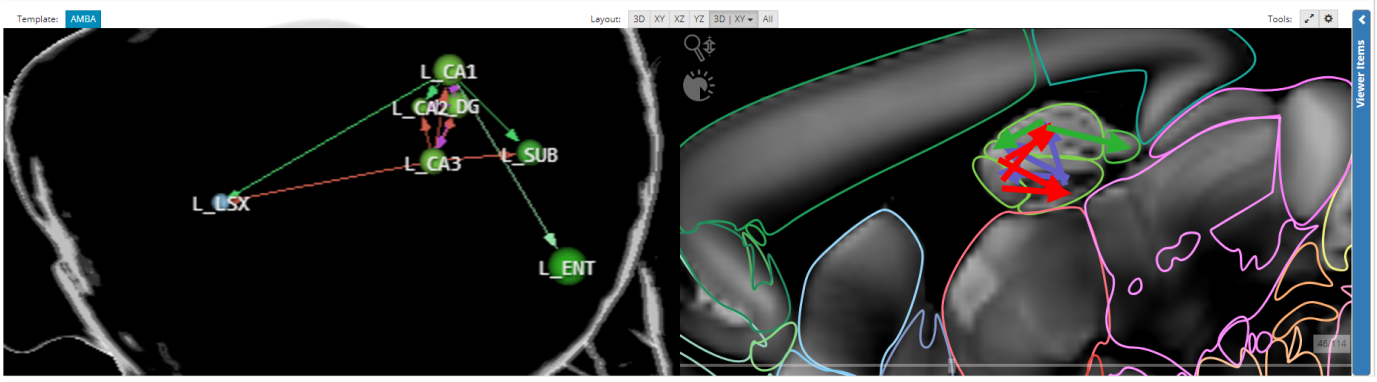
started their investigation by visualizing the spatial gene expression of BDNF in 2D/3D and as *Expression Profile* (Figure 7A) in the *Network Query* tab (W2). Since the circuit has its origin in DG, the user navigates in the *Parcellation Browser*. In the 2D slice view the user sees that there is a high expression level, hence she starts a high-intensity-selection in the *Expression Profile* (Figure 7A).

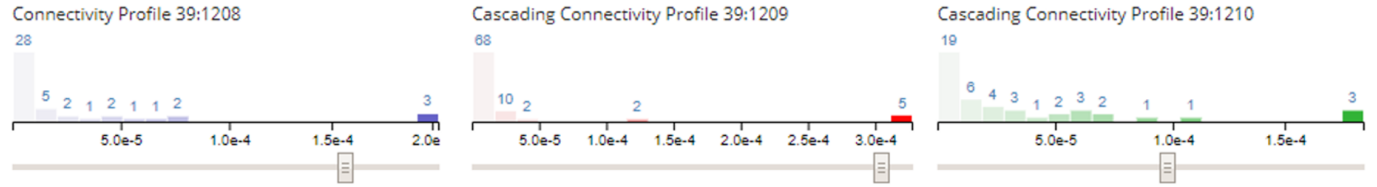
Find connected areas iteratively: By selecting the structural connectivity matrix in the *Query Toolbar* and clicking the *Target Query* button, the user receives the accumulated connectivity instantly as 2D/3D visualization and *Connectivity Profile*. The name of the profile is indented and prefixed with "First-Order" to highlight the iterative procedure (Figure 7A). CA2 and CA3 are the strongest connections (not counting DG, since it is a connection to itself), so they are chosen for the next high-intensity-selection to go further along the circuit (T2). Note that CA1 receives strong input, but this is primarily caused by their spatial closeness and data acquisition technique for structural connectivity [54]. The next *Connectivity Profile* shows CA1 as strongest connection (except for the originating CA2/CA3). By performing a high-intensity-selection on CA1, the results show strong connectivity in the Retrohippocampal Region. After browsing its subregions in the *Parcellation Browser*, the connections to Entorhinal Cortex and Subiculum are revealed (highlighted in Figure 7A bottom)

Visualize the Circuit: Finally, the experts wanted to see the circuit in a 2D-3D *Graph Representation* to give them a spatial context (T3), so that they can use their exploration for presentation and discussion purposes with colleagues (T8). They selected the "Show graph representation" button next to the profiles, which instantly shows them the network graphs (Figure 7B) in the *Network Analysis* tab (W3). After filtering for



A





B

Fig. 7. Case Study 2: A) Gene expression (2D slice view in the black box, selected *VOI* with high-intensity-selection in yellow) and Expression Profile of BDNF (Dentate Gyrus brain-region indicated by a red-arrow) and its first-, second- and third-order *Connectivity Profiles*. B) Graph representation of the connectivity in 2D and 3D (first-order blue, second-order red, third-order green).

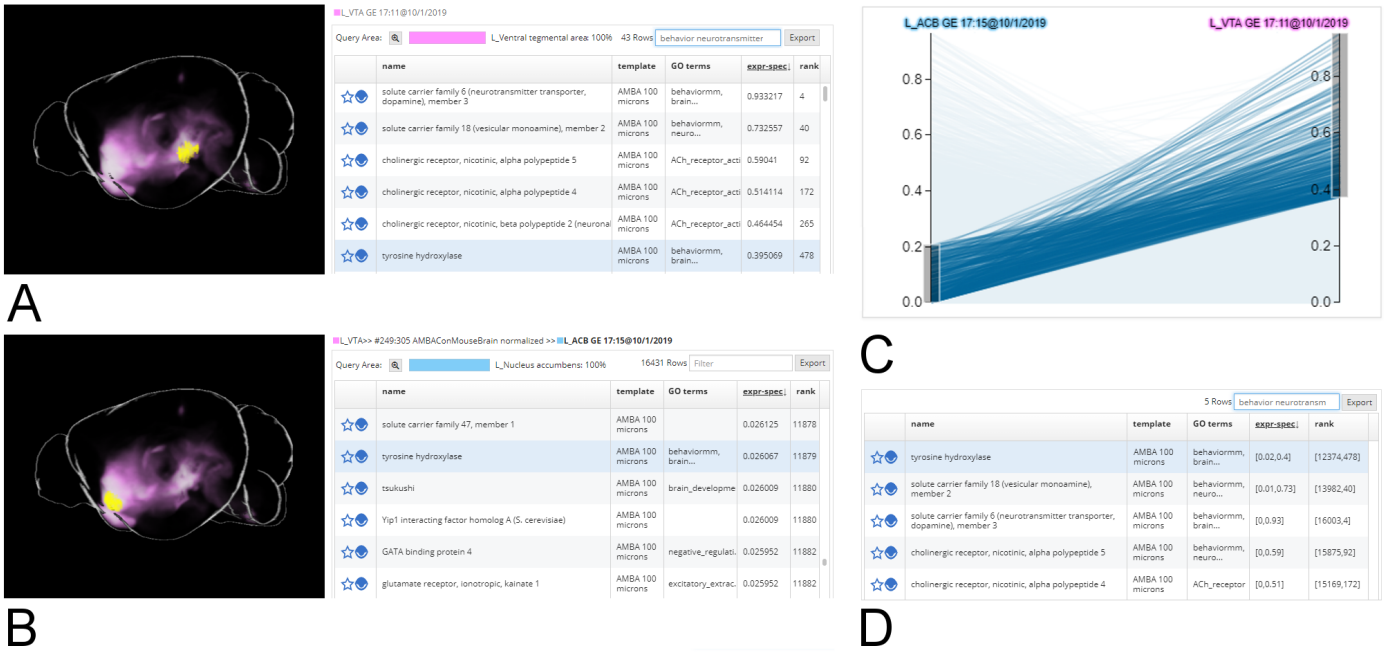


Fig. 8. Case Study 3: A) Outgoing structural connectivity (pink) of VTA (yellow) and genes with the highest mean expression specificity at this region. Genes are filtered by *GO terms behavior* and *neurotransmitter*. B) Strongest targets voxels of VTA within ACB (yellow) and genes with low mean expression specificity at this region. C) The parallel coordinates system shows the selection of genes with high specificity for VTA and low specificity for ACB. D) Genes that are selected by the parallel coordinates system.

the strongest connections, the graph shows only the DG → CA2/CA3 → CA1 → Entorhinal cortex/Subiculum circuit (and a connection to the lateral septal complex LTX, which is not relevant for this case study). Since the *Connectivity Profiles* were automatically marked with different colors, the colored histograms in the *Network Analysis Toolbar*, that are ordered according to their iteration, allow association of graph edges with the connectivity-order originating from DG.

In this case study, we showed the iterative exploration of a 90 GB connectivity matrix at voxel-level, enabled by quantitative information in *Connectivity Profile* and high-intensity *VOI* selection. To our best knowledge, this could be done so far only at region-level connectivity. We provide a video of this study as *Supplementary Video 2*.

8.3. Genetic dissection of connectivity

In this case study, our domain experts sought to evaluate whether the genetic dissection of a well-known relation in the mesolimbic system was feasible. The structural connection from the ventral tegmental area (VTA midbrain) to the nucleus accumbens (ACB basal forebrain) plays a role in the dopamine reward circuit [55]. They sought to examine this connection genetically by querying the gene expression index for highly specific genes in VTA and its structural targets in ACB. The results can be evaluated by exploring them for genes with known functional relations to this circuit (mouse gene ENTREZ ID in brackets): tyrosine hydroxylase (21823) and the dopamine receptors D1, D2, D3 and D5 (13488,13489,13490,13492). Structural connectivity at a 100-micron resolution from *AMBA* is added to the workspace in the *Browse Database* tab (W1) before the case study.

Select volumes of interest for genetic dissection: The entry point for the genetic dissection of the circuit is the VTA brain region. Therefore, the user selects VTA in the *Parcellation Browser* and sets the region as *VOI* (Figure 8A, yellow) for a *Gene Expression Query*.

Find genes that are specific for a volume of interest: A list of genes, ranked by their specificity for VTA is retrieved by executing a "High Expression Query" (T6) from the *Query Toolbar* in the *Gene Expression Query* tab (W4). Tyrosine hydroxylase is ranked 478th of 19479 genes, which puts it into the top 2.5%. Dopamine receptors, which are not specific for VTA ranked mainly around 10.000. This indicates that a *Gene Expression Query* returns meaningful results, but it is rather unrealistic that a user would scroll over 450 genes to find specific genes. In an exploratory setting, a user would use these queries with an a-priory assumption of which genes are relevant for her (i.e. genes related to a certain function/behavior). Therefore, our domain experts filter this list by *GO terms* relevant for the reward circuitry, such as *behavior* and *neurotransmitter*. This puts tyrosine hydroxylase at the 6. place (Figure 8A).

Find genes that are specific for connected areas: To explore the structural targets of VTA, the user selects the structural connectivity matrix in the *Query Toolbar* (*Network Query* tab, W2) and starts a *Target Query*. The targets (Figure 8A, pink) highlight a region in the striatum. In the *Connectivity Profile* the user sees that the ACB is among the strongest connected regions, therefore, she wants to compare its gene expression to VTA. A high-intensity-selection within the ACB sets the voxels with the highest connectivity as *VOI* (Figure 8B, yellow). A "High Expression Query" (T6) in the *Gene Expression Query* tab (W4) reveals gene expression specificity contrary to VTA:

tyrosine hydroxilase has a low specificity (Figure 8B), while the dopamine receptors are ranked on average in the top 1% (after filtering by GO term, the majority of them are in the top 10).

Compare gene expression specificity of brain-regions: By using a parallel coordinates system (Figure 8C) in the *Gene Expression Analysis* tab (W5), the user can search for genes with characteristic specificity in different brain regions (T7), such as ligand-receptor binding (i.e. a gene is expressed in one region, but not in another [45]). This specific case requires expression specificity of the ligand at the projecting site and receptor for this ligand at the target site. To filter for genes with this property, the user selects high specificity for VTA, and low specificity for ACB on the coordinate axes, which show the previously generated gene lists of VTA and ACB. This narrows the selection down to around 440 genes. These are still to many genes to check manually, so the user applies additional filtering by the *GO terms behavior* and *neurotransmitter*. Only 5 genes including tyrosine hydroxylase remain (Figure 8D). This means that without prior knowledge of tyrosine hydroxylase expression, the user would have been able to identify VTA as a candidate source of dopamine in a reward circuit. Similar accounts for dopamine receptors when selecting the opposite relations in gene expression.

This case study shows the genetic dissection of a neuroanatomical connection. It highlights how voxel-level connectvity and genome-wide spatial gene expression can be combined and explored interactively with BrainTrawler. A video of this case study is provided as *Supplementary Video 3*.

9. Discussion and Evaluation

Section 8 showed the potential and relevance of our method in the context of neuroscientific research, but there is still room for improvements that can be targeted in future work. Therefore, during the case studies, we conducted discussions with 2 domain experts (which are our co-authors Wulf Haubensak and Joanna Kaczanowska, who were also involved in the tool design). Furthermore, we reproduced the case studies with 3 other domain experts (a technician, a PhD student and a postdoc from the field of neuroscience/circuit dissection with affiliations to the University of Vienna), who have seen the tool before (but not in detail). The user experience was recorded in informal but focused interviews. The following covers the feedback derived from these discussions/interviews.

Visual Design and User Interactions: All domain experts stated that the framework was well designed and meaningful. Especially the visual queries were of great interest for them. *VOI* selections were reasonable to perform and intuitively understood. The brush-selection was considered to be the most intuitive one, while the high-intensity selection was perceived as most accurate one. Also the region-selection had its application in a first, coarse probing of the data. Color schemes and names of the *Connectivity Volumes* were seen as helpful for differentiating different connectivities based on their *VOI*.

To allow a quantitative evaluation on brain-region level, the results are shown as region-wise profile, presenting the mean-connectivity for every brain region, as well as for all its subregions in a bar chart. This directly highlights strongly connected subregions, which would otherwise be missed in case the domain-experts operated at a higher brain region level.

The region-level profile of connectivity further allows for *VOI* selection for transitive connectivity. Although it would generally be possible for a user to manually select strongly connected voxels in the 2D slice view in order to start another target/source query, our domain experts preferred to query for transitive connectivity within brain regions of their interest (i.e. they were not interested in retrieving the connectivity of all target/source sites, but only from sites within brain regions of interest).

Connectivity Query results can be abstracted as 2D and 3D network graphs, which reduce the complexity of voxel-level while incorporating neuroanatomical context. Multiple graphs (i.e. region-level graph representation of connectivity query results) can be rendered with parallel arrows/lines, or combined by multiplication. Our domain experts found this the most intuitive method, since one can directly see the similarity/contrast of two edges between two nodes. They understood that the *Network Analysis Toolbar* provides information about the edge weights as well as a filtering for weak edges. Nevertheless, an additional legend about the meaning of the encodings would be considered as helpful, e.g. it was not clear that the sphere size encodes region size. Also, edge bundling/routing algorithms were considered to produce more visual appealing figures for scientific publications. Otherwise, they felt familiar with this type of visualization.

Filterable Gene Lists for *Gene Expression Query* results provide information in addition to gene expression specificity. According to our domain experts, filtering by *GO terms* was significantly helpful when browsing the lists, for they give a better overview of the functional context. Several domain experts noted, that without filtering for *GO terms*, the list could not be used in an exploratory way (i.e. one need to know for which genes one is looking for). Further filtering in the parallel coordinate system was supported by the colors/names. It was considered to be useful for spatially differentiating the queries' anatomical origin.

Usability and Potential: The overall workflow provided by *BrainTrawler* was intuitive to our domain experts in general. It was noted that tooltips and legends for the network graphs could make it more clearer, especially if one does not use *BrainTrawler* very often. An export function for *Connectivity Profiles* and network graphs as csv file was a common request among the participants, so they could use the data for different analyses in other tools. They were explicitly asked if they would consider including it into their own experimental workflow. Their answer was yes. They would mainly use it for hypothesis building, and thus as an entry point for their experiments. It would save them time in contrast to manually searching online databases instead. Initial hypothesis building is part of the experimental workflow of all participants, so there was no difference on how their background (technician, PhD student, postdoc) would effect their work with *BrainTrawler*. Asked for the potential of this work, our domain experts stated that this framework could be either used to present large-scale

resources online, or for refinement of functional neuroanatomy.

User Interface Limitations: In general, there is no limit how much data can be loaded into the workspace, but there are several limitations regarding the user interface. The 3D visualizations support 4 simultaneously rendered volumes (i.e. gene expression data or *Connectivity Volumes*), but even more than 3 were perceived as too obstructive. Similar limitations apply to the 2D slice view, although the overlap can be visualized for more volumes. To analyze expression data of more genes (e.g. sets of genes related to a certain behavior or brain function) abstraction methods would be needed such as rendering the mean or maximum gene expression for each voxel. For the 3D/2D network visualization, including more than 3 networks leads to too many parallel edges. In this case, the overlap visualization (multiplication of all edges) can be used to reduce the total amount of edges. Comparing more than 3 *Connectivity and Expression Profiles* was perceived as difficult when they are shown in an overview below each other. For gene lists in the *Gene Expression Analysis*, the parallel coordinates system does not scale well to more than 15 queries, for there is not enough space in the user interface. Nevertheless, none of these issues were considered as limitations for our case studies or potential further future projects.

Performance: *Connectivity* and *Gene Expression Queries* were instantly executed without an inconvenient delay for the user (results in $\leq$ 2 seconds) during the case studies. In Ganglberger et al. [7] we showed, that on SSD, *Connectivity Queries* can be executed in less than one second for brain areas involving about 1% of the brain, and less than four seconds for larger areas up to the whole brain. On HDD the queries are approximately three times slower. We performed these tests again for *Split-Aggregation Queries* without major performance losses. *Gene Expression Queries* were notably slower for larger brain areas (linear with *VOI* size), since the implementation of the spatial index ([9]) did not utilize hierarchical structures similar to [7].

10. Conclusion

In this paper we present *BrainTrawler*, a novel web-framework for analyzing and fusing heterogeneous neurobiological data of different types, modalities and scale via their spatial context. The framework integrates data from different large-scale brain initiatives as well as user generated data. It incorporates a hierarchically organized data structure, which enables real-time querying and aggregating of huge brain network connectivity of different scales and resolutions in a common standard brain space. In combination with spatial indexing of vast gene expression collections, for the first time ever it is possible to genetically dissect large-scale brain networks on a voxel-level in real-time. This enables neuroscientists to explore the genetic and functional characteristics of microcircuits without time-consuming manual data-aggregation and literature-research. Furthermore, they can share their results as link, with visualizations that neuroscientists are familiar with, to support collaboration and data provenance in a scientific environment. Finally, the case studies conducted with domain experts showed biological validity by reproducing findings of known microcircuits that are subject to current research.

For the future, we aim to extend this framework in a holistic way. Therefore, it should not only allow to access the data, but also include the import of user-generated data and preprocessing. One could further improve network visualization by incorporating edge bundling/routing algorithms or sophisticated 2D graph layouts. Quantitative analysis of networks would benefit from computing network statistics and more detailed graphs legends. Furthermore, we want to position *BrainTrawler* as tool for the presentation of large-scale resources of neuroimaging initiatives.

11. Information Sharing Statement

BrainTrawler is available upon individual request to the corresponding author, Florian Ganglberger or Katja Bühler.

12. Acknowledgments

This work is a result of a joint IMP VRVis project supported by Grant 852936 of the Austrian FFG Funding Agency. W. H. was supported by a grant from the European Community's Seventh Framwork Programme (FP/2007-2013) / ERC grant agreement no. 311701, the Research Institute of Molecular Pathology (IMP), Boehringer Ingelheim and the Austrian Research Promotion Agency (FFG). VRVis is funded by BMVIT, BMDW, Styria, SFG and Vienna Business Agency in the scope of COMET—Competence Centers for Excellent Technologies (854174) which is managed by FFG. Special thanks to Dominic Kargl, Lukasz Piszczek and Andreea Constantinescu (all three are associated with the University of Vienna/IMP) for his participation in the Case Study.

References

[1] alleninstitute. Allen institute. 2019. URL: `https://www.alleninstitute.org/`.

[2] humanbrainproject. Human brain project. 2019. URL: `https://www.humanbrainproject.eu`.

[3] Van Essen, DC, Smith, SM, Barch, DM, Behrens, TE, Yacoub, E, Ugurbil, K, et al. The wu-minn human connectome project: an overview. Neuroimage 2013;80:62–79.

[4] ming Poo, M, lin Du, J, Ip, NY, Xiong, ZQ, Xu, B, Tan, T. China Brain Project: Basic Neuroscience, Brain Diseases, and Brain-Inspired Computing. Neuron 2016;.

[5] Kim, J, Zhang, X, Muralidhar, S, LeBlanc, SA, Tonegawa, S. Basolateral to central amygdala neural circuits for appetitive behaviors. Neuron 2017;93(6):1464–1479.

[6] Ganglberger, F, Swoboda, N, Frauenstein, L, Kaczanowska, J, Haubensak, W, Bühler, K. Iterative exploration of big brain network data. In: VCBM 18: Eurographics Workshop on Visual Computing for Biology and Medicine, Granada, Spain, September 20-21, 2018. 2018, p. 77–87.

[7] Ganglberger, F, Kaczanowska, J, Haubensak, W, Buehler, K. A data structure for real-time aggregation queries of big brain networks. bioRxiv 2018;URL: `https://doi.org/10.1101/346338`.

[8] Lein, ES, Hawrylycz, MJ, Ao, N, Ayres, M, Bensinger, A, Bernard, A, et al. Genome-wide atlas of gene expression in the adult mouse brain. Nature 2007;445(7124):168–176.

[9] Bruckner, S, Soltészová, V, Gröller, ME, HladÅvka, J, Bühler, K, Yu, JY, et al. BrainGazer–visual queries for neurobiology research. IEEE transactions on visualization and computer graphics 2009;15(6):1497–504.

[10] Arganda-Carreras, I, Manoliu, T, Mazuras, N, Schulze, F, Iglesias, JE, Bhler, K, et al. A statistically representative atlas for mapping neuronal circuits in the drosophila adult brain. Frontiers in Neuroinformatics 2018;12:13.

[11] Muenzing, SE, Strauch, M, Truman, JW, Bühler, K, Thum, AS, Merhof, D. larvalign: Aligning Gene Expression Patterns from the Larval Brain of Drosophila melanogaster. Neuroinformatics 2018;.

[12] Lau, C, Ng, L, Thompson, C, Pathak, S, Kuan, L, Jones, A, et al. Exploration and visualization of gene expression with neuroanatomy in the adult mouse brain. BMC Bioinformatics 2008;9(1):153.

[13] Feng, D, Lau, C, Ng, L, Li, Y, Kuan, L, Sunkin, SM, et al. Exploration and visualization of connectivity in the adult mouse brain. Methods 2015;73:90–97.

[14] Bezgin, G, Reid, AT, Schubert, D, Kötter, R. Matching spatial with ontological brain regions using java tools for visualization, database access, and integrated data analysis. Neuroinformatics 2009;7(1):7–22.

[15] Margulies, DS, Böttger, J, Watanabe, A, Gorgolewski, KJ. Visualizing the human connectome. NeuroImage 2013;80:445–461.

[16] Huisman, S, vanLew, B, Mahfouz, A, Pezzotti, N, Höllt, T, Michielsen, L, et al. BrainScope: interactive visual exploration of the spatial and temporal human brain transcriptome. Nucleic Acids Research 2017;.

[17] Schneider-Mizell, CM, Gerhard, S, Longair, M, Kazimiers, T, Li, F, Zwart, MF, et al. Quantitative neuroanatomy for connectomics in Drosophila. eLife 2016;5.

[18] Jiang, H, Van Zijl, PCM, Kim, J, Pearlson, GD, Mori, S. DtiStudio: Resource program for diffusion tensor computation and fiber bundle tracking. Computer Methods and Programs in Biomedicine 2006;81(2):106–116.

[19] LaPlante, RA, Douw, L, Tang, W, Stufflebeam, SM. The Connectome Visualization Utility: Software for Visualization of Human Brain Networks. PLoS ONE 2014;9(12).

[20] Mahfouz, A, Huisman, SMH, Lelieveldt, BPF, Reinders, MJT. Brain transcriptome atlases: a computational perspective. Brain Structure and Function 2016;:1–24.

[21] Xia, MR, Wang, JH, He, Y. BrainNet Viewer: a network visualization tool for human brain connectomics. PloS one 2013;8(7).

[22] Xia, M, He, Y. Functional Connectomics from a Big Data Perspective. NeuroImage 2017;.

[23] Scholtens, LH, Schmidt, R, De Reus, MA, Van Den Heuvel, MP. Linking Macroscale Graph Analytical Organization to Microscale Neuroarchitectonics in the Macaque Connectome. The Journal of Neuroscience 2014;.

[24] Poldrack, RA, Laumann, TO, Koyejo, O, Gregory, B, Hover, A, Chen, MYea. Long-term neural and physiological phenotyping of a single human. Nature Communications 2015;6:8885.

[25] Huntenburg, JM, Bazin, PL, Goulas, A, Tardif, CL, Villringer, A, Margulies, DS. A Systematic Relationship Between Functional Connectivity and Intracortical Myelin in the Human Cerebral Cortex. Cerebral Cortex 2017;:1–17.

[26] Lin, CY, Tsai, KL, Wang, SC, Hsieh, CH, Chang, HM, Chiang, AS. The neuron navigator: Exploring the information pathway through the neural maze. In: IEEE Pacific Visualization Symposium 2011, PacificVis 2011 - Proceedings. ISBN 9781612849324; 2011,.

[27] Saalfeld, S, Cardona, A, Hartenstein, V, Tomančák, P. CATMAID: Collaborative annotation toolkit for massive amounts of image data. Bioinformatics 2009;.

[28] Beyer, J, Al-Awami, A, Kasthuri, N, Lichtman, J, Pfister, H, Hadwiger, M. Connectomeexplorer: Query-guided visual analysis of large volumetric neuroscience data. IEEE Transactions on Visualization and Computer Graphics (Proceedings IEEE SciVis 2013) 2013;19(12):2868–2877.

[29] Sherbondy, A, Akers, D, Mackenzie, R, Dougherty, R, Wandell, B. Exploring connectivity of the brain's white matter with dynamic queries. In: IEEE Transactions on Visualization and Computer Graphics. ISBN 1077-2626 (Print); 2005,.

[30] Al-Awami, AK, Beyer, J, Strobelt, H, Kasthuri, N, Lichtman, JW, Pfister, H, et al. Neurolines: A subway map metaphor for visualizing nanoscale neuronal connectivity. IEEE Transactions on Visualization and Computer Graphics 2014;20(12):2369–2378.

[31] Nowke, C, Schmidt, M, van Albada, SJ, Eppler, JM, Bakker, R, Diesrnann, M, et al. VisNEST - Interactive analysis of neural activity data. Biological Data Visualization (BioVis), 2013 IEEE Symposium on 2013;:65–72.

[32] Schmitt, O, Eipert, P. neuroVIISAS: Approaching Multiscale Simulation of the Rat Connectome. Neuroinformatics 2012;10(3):243–267.

[33] Li, H, Fang, S, Goni, J, Contreras, JA, Liang, Y. Integrated Visualization of Human Brain Connectome Data. Lecture Notes in Computer 2015;9250:295–305.

[34] Balzer, M, Deussen, O. Level-of-detail visualization of clustered graph layouts. Asia-Pacific Symposium on Visualisation 2007, Proceedings 2007;:133–140.

[35] Boettger, J, Schaefer, A, Lohmann, G, Villringer, A, Margulies, DS. Three-Dimensional Mean-Shift Edge Bundling for the Visualization of Functional Connectivity in the Brain. IEEE Transactions on Visualization and Computer Graphics 2014;20(3):471–480.

[36] Murugesan, S, Bouchard, K, Brown, JA, Hamann, B, Seeley, WW, Trujillo, A, et al. Brain Modulyzer : Interactive Visual Analysis of Functional Brain Connectivity. IEEE/ACM transactions on computational biology and bioinformatics 2016;:1–14.

[37] Keiriz, JJG, Zhan, L, Ajilore, O, Leow, AD, Forbes, AG. NeuroCave: A web-based immersive visualization platform for exploring connectome datasets. Network Neuroscience 2018;2(3):344–361.

[38] Sorger, J. neuroMap - Interactive Graph-Visualization of the Fruit Fly ' s Neural Circuit. In: Biological Data Visualization (BioVis), 2013 IEEE Symposium on. 2013,.

[39] Ji, C, Maurits, NM, Roerdink, JBTM. Data-driven visualization of multichannel EEG coherence networks based on community structure analysis. Applied Network Science 2018;3(1):41.

[40] Alper, B, Bach, B, Henry Riche, N, Isenberg, T, Fekete, JD. Weighted Graph Comparison Techniques for Brain Connectivity Analysis. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '13; New York, NY, USA: ACM. ISBN 978-1-4503-1899-0; 2013, p. 483–492.

[41] de Ridder, M, Klein, K, Yang, J, Yang, P, Lagopoulos, J, Hickie, I, et al. An Uncertainty Visual Analytics Framework for fMRI Functional Connectivity. Neuroinformatics 2018;.

[42] Bach, B, Henry-Riche, N, Dwyer, T, Madhyastha, T, Fekete, JD, Grabowski, T. Small MultiPiles: Piling Time to Explore Temporal Patterns in Dynamic Networks. Computer Graphics Forum 2015;34(3):31–40.

[43] Haubensak, W, Kunwar, PS, Cai, H, Ciocchi, S, Wall, NR, Ponnusamy, R, et al. Genetic dissection of an amygdala microcircuit that gates conditioned fear. Nature 2010;468(7321):270–276.

[44] Oh, SW, Harris, JA, Ng, L, Winslow, B, Cain, N, Mihalas, S, et al. A mesoscale connectome of the mouse brain. Nature 2014;508(7495):207–214.

[45] Young, LJ, Wang, Z. The neurobiology of pair bonding. Nature neuroscience 2004;7(10):1048–1054.

[46] Genç, E, Schölvinck, ML, Bergmann, J, Singer, W, Kohler, A. Functional connectivity patterns of visual cortex reflect its anatomical organization. Cerebral Cortex 2015;:bhv175.

[47] AMBAGeneExpressionWhitepaper. Allen mouse brain atlas gene expression whitepapers. 2018. URL: `http://help.brain-map.org/display/mousebrain/Allen+Mouse+Brain+Atlas`.

[48] Zalesky, A, Fornito, A, Bullmore, ET. Network-based statistic: Identifying differences in brain networks. NeuroImage 2010;53(4):1197–1207.

[49] Richiardi, J, Altmann, A. Correlated gene expression supports synchronous activity in brain networks. Science 2015;348(6240):11–14.

[50] Bassett, DS, Sporns, O. Network neuroscience. Nature Neuroscience 2017;.

[51] Carbon, S, Dietze, H, Lewis, SE, Mungall, CJ, Munoz-Torres, MC, Basu, S, et al. Expansion of the gene ontology knowledgebase and resources: The gene ontology consortium. Nucleic Acids Research 2017;.

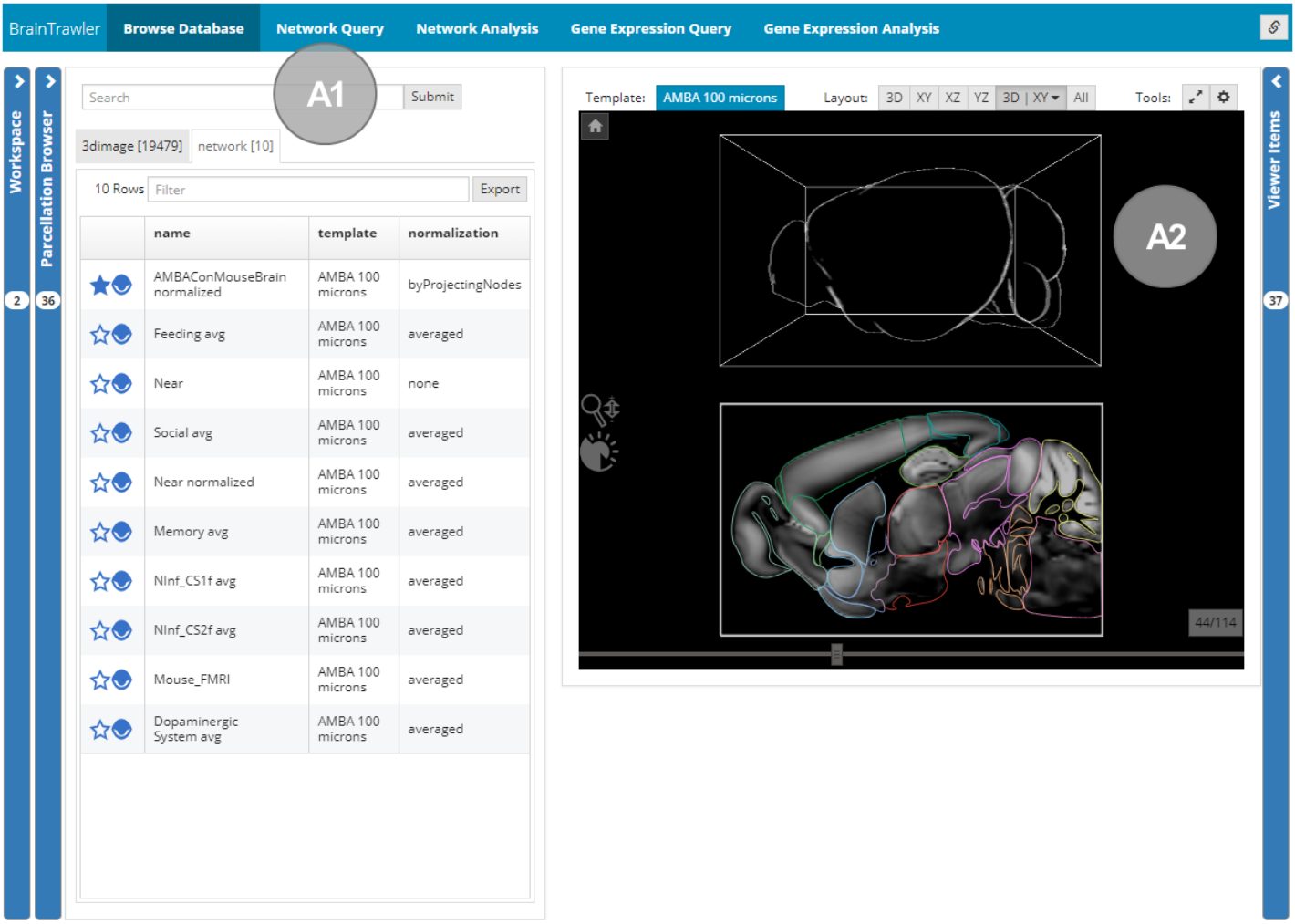
[52] Baribeau, DA, Anagnostou, E. Oxytocin and vasopressin: Linking pituitary neuropeptides and their receptors to social neurocircuits. 2015.

[53] Leal, G, Bramham, C, Duarte, C. Chapter eight - bdnf and hippocampal synaptic plasticity. In: Litwack, G, editor. Neurotrophins; vol. 104 of *Vitamins and Hormones*. Academic Press; 2017, p. 153 – 195.

[54] AMBAConnectivityWhitepaper. Allen mouse brain atlas connectivity whitepapers. 2018. URL: `http://help.brain-map.org/display/mouseconnectivity/Documentation`.

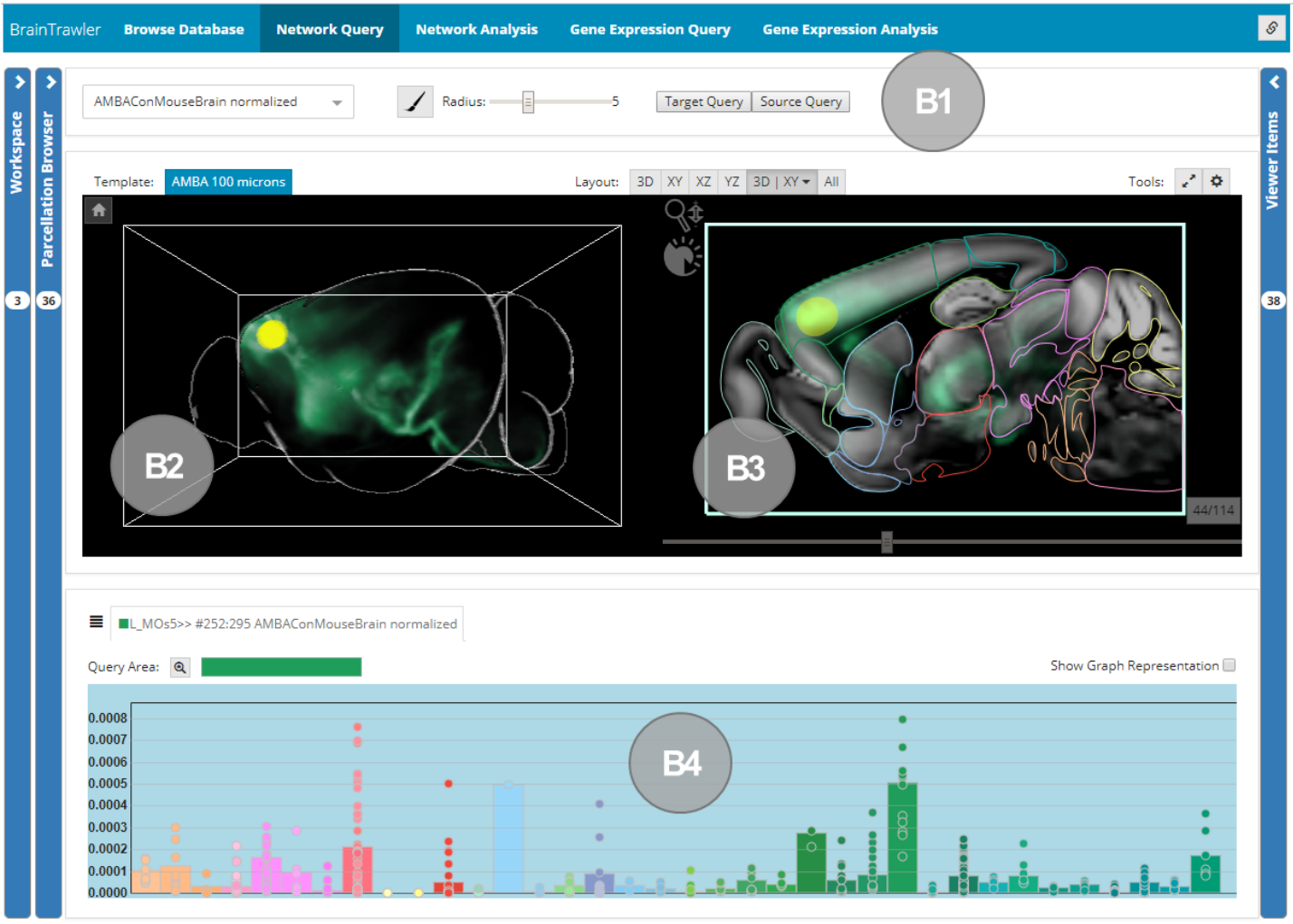
[55] Malenka, R, Nestler, E, Hyman, S. Chapter 6: Widely Projecting Systems: Monoamines, Acetylcholine, and Orexin. In: Molecular Neuropharmacology: A Foundation for Clinical Neuroscience. (3rd edn). 2009,.

Supplement



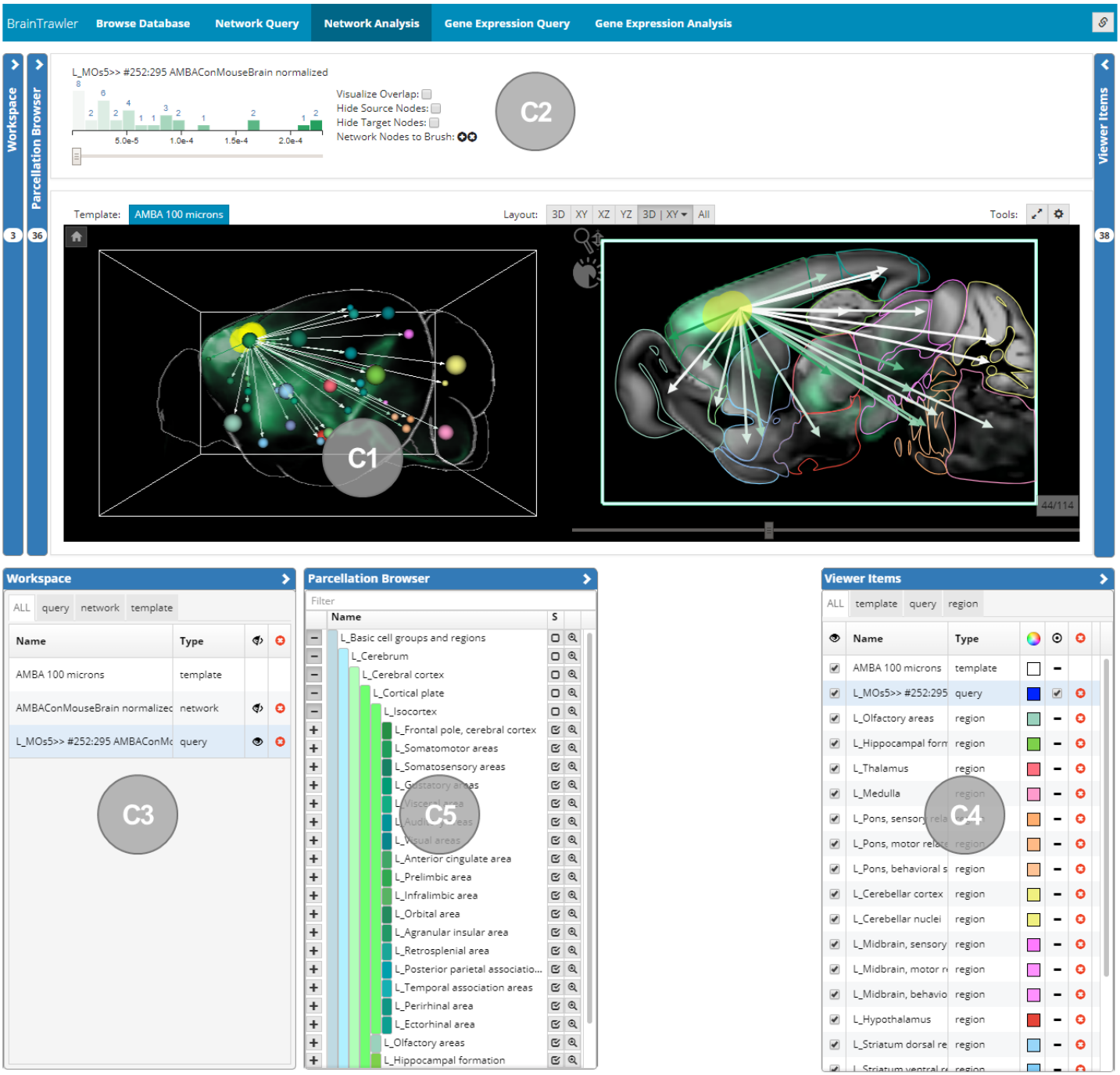
A

Supplementary Figure 1: **A)** Browse Database tab: The user can browse a collection of 3D gene expression and connectivity data via a text search (A1) and visualize the data directly in the 3D view (A2) or add it to the workspace (C3).



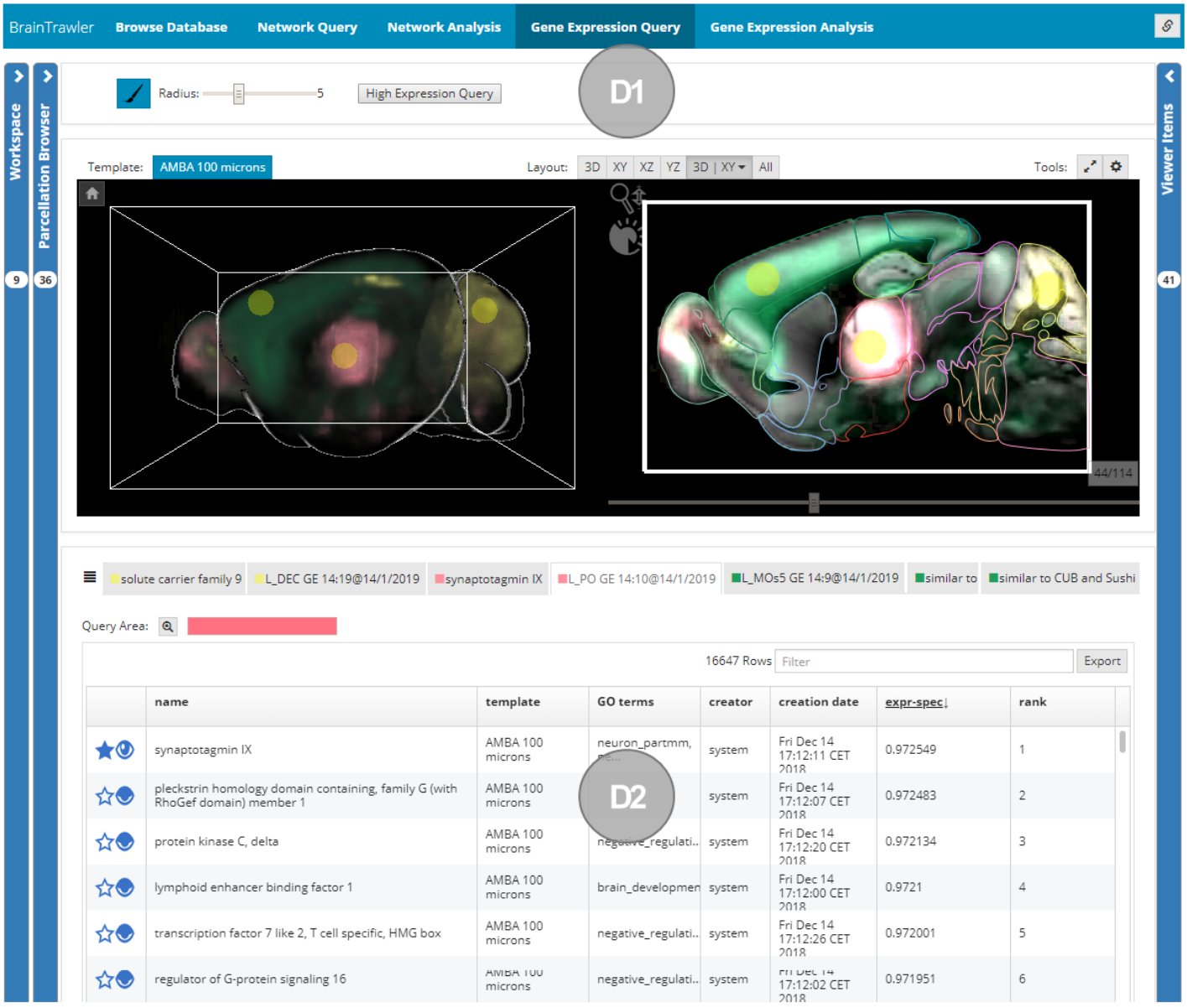
B

Supplementary Figure 2: **B)** Network Query tab: The user can execute target/source queries from the Query Toolbar (B1) on connectivity matrices that have been added to the workspace. Gene expression/Connectivity is rendered volumetrically in a 3D view (B2), in a 2D slice view (B3) and as region-wise quantitative representation (Expression/Connectivity Profile) (B4).



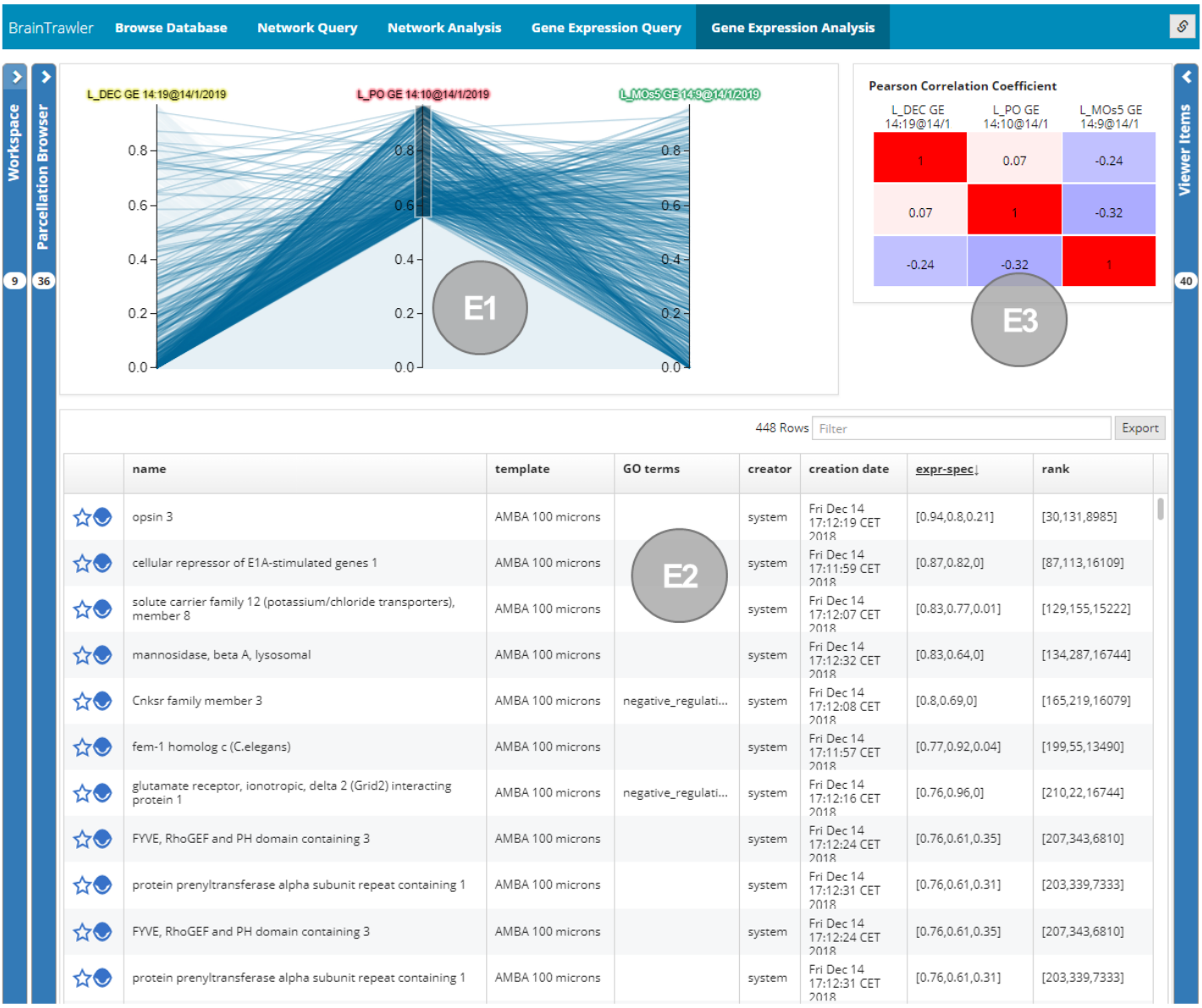
C

Supplementary Figure 3: **C)** Network Analysis tab: The user can explore networks or connectivities on a region level in a ball/stick representation (C1), controlled by the Network Analysis Toolbar showing histograms of the visualized graphs and threshold sliders for edge filtering (C2). The workspace manages which items (3D volumetric data, networks, query results) are shown in each tab (C3) while the Viewer Items List (C4) controls their appearance (e.g. color). The Parcellation Browser (C5) shows a hierarchical (anatomical) brain parcellation and is used to change the parcellation level of visualized graphs and profiles.



D

Supplementary Figure 4: **D)** Gene Expression Query tab: Via the Query Toolbar (D1), *Gene Expression Queries* can be executed. Resulting gene lists are shown in a table below (D2).



E

Supplementary Figure 5: **E)** Gene Expression Analysis tab: Gene lists can be compared in a parallel coordinate system (E1). Its selection is shown in a table (E2), its correlation in a heatmap (E3).

List of Figures

1.1 Effects of a gene in the mouse brain. Spatial gene expression of the gene PKC-Delta is shown in blue: (A) Primary expression sites: Effect on a brain region of interest where the gene is expressed (red circle). (B) Second-order effects: Projections from/to the brain region of interest (red circles). (C) Higher-order effects: Projections from/to the second-order effects (red circles). 4

1.2 Predicted functional neuroanatomical map of social bonding behavior in the mouse brain. Red color indicates where the areas are and the transparency shows how strong these areas are associated. 5

1.3 Aggregation Query: The aggregated outgoing connectivity of a brain region (red circle) can be computed by accumulating the outgoing connectivity of every voxel within the brain region in a connectivity matrix. The result is the sum of all outgoing connections from the brain region to the rest of the brain (green cloud). 6

1.4 Exemplary sub-network in the mouse brain on different levels of anatomical abstraction: The green point cloud represents the outgoing structural connectivity of the yellow area. The 3D node-link diagram represents this connectivity at a brain region-level. Connectivity is outgoing from the left motor cortex (L_MO). Spheres identify the center of brain regions, while the size encodes the region size. The intensity of the arrows indicates the connection strength (i.e., how many voxels in this region have a connection from L_MO). 7

1.5 The papers of this thesis categorized by their level of interaction with the underlying data and the interactivity of the visualization. The figures indicate the papers' key contributions. 8

2.1 Determination of relationships between brain regions/networks, behavior/functions and genes. Black arrows: Alterations in the genome of model organisms (e.g., mouse) can be studied to identify the function of specific genes and their influence on the organism's behavior. Via fMRI, electrophysiology, or optophysiology, these effects can be linked to brain regions or networks. Red arrows: From literature, genetic databases, or genome-wide association studies, collections of genes can be associated to behavior/function. By mapping these genes to the brain via spatial brain data of big brain initiatives, the effect of behavior/function/genes on the brain can be studied computationally. . . 12

2.2 Imaging Data: (A) Region-level data: One value for every brain region. (B) Voxel-level data: Volumetric data, one value for every voxel. 15

2.3 Anatomical Data: (A) Anatomical reference template showing the anatomical structure of a mouse brain (2D sagittal slice). (B) Part of a hierarchical neuroanatomical ontology. (C) Brain parcellation that corresponds to the brain ontology (2D sagittal slice). 16

2.4 Connectivity data types: (A) Sparse directed network (structural connectivity) (B) Dense undirected network (functional connectivity, genetic connectivity) (C) Dense directed network (effective connectivity). 18

2.5 3D network visualization. (A) Region-level network visualization via 3D node-link diagram in the BrainNet Viewer ([XWH13], Figure 5). Colors represent different subnetworks, node size represents node strength (number of connections of a node). Orange edges represent long distance connections. (B) Voxel-level network visualization of functional connectivity (orange: negative correlation, green: positive correlation) with edge-bundling ([BSJ$^+$14], Figure 13). 20

2.6 Node-link diagram for network visualization with anatomical context. BrainModulizer ([MBB$^+$16], Figure 8) shows a 2D graph, split into four brain modules (red, blue, green, purple) rendered with a force-directed layout (B) next to its 3D anatomical context (A, colors correspond to nodes). 21

2.7 A network graph in 3D anatomical space (left) showing four brain modules (orange, blue, green and red) visualized in NeuroCave ([KZA$^+$18], Figure 5). The right panel shows the same graph, visualized in a topological space (connectivity-driven). 22

2.8 Anatomical layouts: (A) Anatomical layout of a fruit fly brain (drosophila). Blue regions are major brain regions of the fruit fly brain. Network nodes are arranged within these regions (NeuroMap [Sor13], Figure 6). (B) Planar projection of 3D electrode locations (nodes) on the human skull. Functional connectivity of a network module is visualized between these locations (Ji et al. [JMR18], Figure 7). 23

2.9 Examples for network comparison: (A) Small multiples of four functional networks in a circular layout (brain regions) ([dRKY+18], Figure 3). (B) 125 snapshots of a dynamic network (i.e., 125 states of a network at different time points) mapped to a two dimensional space via PCA ([vdEHBvW16], Figure 8). Stable states (clusters) are highlighted with blue circles (gray graphs show a representative snapshot). 25

2.10 Visualization of spatial gene expression data: (A) Mapping of 3000 genes to a two dimensional space via t-SNE of their gene expression ([HvM+17], Figure 3). Expression of representative genes of several clusters are shown in 2D slice views of the human brain. (B) Spatial gene expression of the gene Man 1a via glyphs, overlaid with anatomical context of the mouse brain ([LNT+08], Figure 4a and c). 27

2.11 3D segmented dendrite (red) and a connected axon(blue) (left) mapped to an abstract representation similar to a "subway map" (right) with NeuroLines ([AABS+14], Figure 1). 28

2.12 3D rendering of segmented neuronal structures in the fruit fly's brain ([SMB+17], Figure 2). 29

2.13 3D visualization of neuronal structures and connections. Connections emerging from a specific brain area (secondary motor area) visualized as tubular trajectories. Color codes indicate the brain regions passed by the connections ([FLN+15], Figure 1). 30

3.1 Concept of GWCA: From genetic meta data to functional neuroanatomical maps 32

3.2 Functional maps generated with GWCA from functional gene sets associated with the central amygdala circuitry, dopaminergic signaling, feeding, hypothalamic circuitry, fear memory consolidation, panic disorder, learning in a stressful context, social bonding, and synaptic plasticity (left to right, top to bottom). The colors emphasize the versatile nature of the gene sets and have no specific meaning. 33

3.3 Comparison of functional maps and fMRI of a mouse brain. (A) A slice of a functional map related to pain. Color indicates how significant a voxel is associated to brain function/behavior. (B) A slice of an fMRI image related during pain perception. Color indicates a BOLD (blood oxygen-level dependent) signal change between control and pain. The olfactory bulb (front part of the mouse brain) was not available, since it has not been imaged during fMRI. 34

3.4 Concepts for a visualization of brain networks within neuroanatomical context. (A) Schematic of brain reward circuitry in a mouse brain by Russo et al., Figure 1 [RN13]. (B) Sagittal projection of the mouse brain. Clusters of the functional maps are colored differently, the short names of the clusters' major anatomical regions are listed in the legend. 35

3.5 Different design stages for a visualization of brain networks with spatial, neuroanatomical context: (A) Sagittal projections of six clusters of functional maps. Voxel color represents how significantly the voxel is associated with a specific behavior/function. Colored circles show the relative number of voxels in the respective brain region. Arrows indicate the structural connectivity in-between. (B) Sagittal projections of five clusters of functional maps, similarly to A. Labels show the major brain regions of each cluster. 36
3.6 A Hilbert curve through a mouse brain volume. Along this curve, locality is preserved. 37
3.7 Schema of a hierarchical mapping: The lowest level represents a voxel-level reference space, while higher levels comprise brain regions. 38
3.8 Network visualization on two different hierarchical levels: The top panel shows the region L_Pons, while the bottom panel shows the sub-regions of L_Pons. 40
3.9 Relating gene expression data to structural connectivity: (A) Visualization of gene expression of the gene PKC-delta (cyan) in a sagittal slice of the mouse brain (anatomical brain regions are outlined). (B) Selection of voxels with high gene expression (yellow) within the Striatum-like amygdala nuclei region (blue outline). (C) Outgoing structural connectivity of the selection (light green). (D) Outgoing structural connectivity of the selection in 3D. (E) *Connectivity Profile* shows the mean connectivity of brain regions as bar chart. Circles represent the connectivity for anatomical sub-regions. 42
3.10 Concept of a *Split Aggregation Query* 43
3.11 Voxel-level connectivity of Figure 3.9 on the region-level: (A) Outgoing structural connectivity of the selection on region-level. (B) Outgoing structural connectivity of the selection on region-level, represented as 3D node-link diagram. The center of the spheres indicate the regions center-of-mass, the size encodes the region size. (C) Switching to a different hierarchy level, the “green regions“ are collapsed, while the “blue regions“ are extended. (D) Outgoing structural connectivity on a different hierarchy level where the region of the volume of interest is split into sub-regions (yellow circles). . 43
3.12 Connectivity comparison on a region-level: (A) Quantitative connectivity comparison of fMRI connectivity and structural connectivity with *Connectivity Profiles*. The bars show the mean connectivity within brain regions. Circles represent sub-regions of the respective region/bar. (B) Connectivity comparison visualizing fMRI as blue arrows and structural connectivity in green. Region colors correspond to the *Connectivity Profile*. (C) Overlap (product) of both connectivities visualized with red arrows. 44

Bibliography

[AABS+14] Ali K. Al-Awami, Johanna Beyer, Hendrik Strobelt, Narayanan Kasthuri, Jeff W. Lichtman, Hanspeter Pfister, and Markus Hadwiger. NeuroLines: A subway map metaphor for visualizing nanoscale neuronal connectivity. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2369–2378, 2014.

[ABH+13] Basak Alper, Benjamin Bach, Nathalie Henry Riche, Tobias Isenberg, and Jean-Daniel Fekete. Weighted Graph Comparison Techniques for Brain Connectivity Analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 483–492, New York, NY, USA, 2013. ACM.

[all19] Allen institute, https://www.alleninstitute.org/, 2019.

[ASO+16] Diego A Angulo, Cyril Schneider, James H Oliver, Nathalie Charpak, and Jose T Hernandez. A Multi-facetted Visual Analytics Tool for Exploratory Analysis of Human Brain and Function Datasets. *Frontiers in Neuroinformatics*, 10:36, 2016.

[BAAK+13] Johanna Beyer, Ali Al-Awami, Narayanan Kasthuri, Jeff W. Lichtman, Hanspeter Pfister, and Markus Hadwiger. ConnectomeExplorer: Query-guided visual analysis of large volumetric neuroscience data. *IEEE Transactions on Visualization and Computer Graphics*, 2013.

[BB17] Richard F. Betzel and Danielle S. Bassett. Multi-scale brain networks. *NeuroImage*, 2017.

[BHRD+15] B Bach, N Henry-Riche, T Dwyer, T Madhyastha, J-D. Fekete, and T Grabowski. Small MultiPiles: Piling Time to Explore Temporal Patterns in Dynamic Networks. *Computer Graphics Forum*, 34(3):31–40, 2015.

[BJV12] Ian Bowman, Shantanu H Joshi, and John D Van Horn. Visual systems for interactive exploration and mining of large-scale neuroimaging data archives. *Frontiers in neuroinformatics*, 6:11, 2012.

[bra11] Braingl, code.google.com/p/braingl, 2011.

[bra19] Brain* software framework, https://braingazer.org/, 2019.

[BRSK09] Gleb Bezgin, Andrew T. Reid, Dirk Schubert, and Rolf Kötter. Matching spatial with ontological brain regions using java tools for visualization, database access, and integrated data analysis. *Neuroinformatics*, 7(1):7–22, 2009.

[BS17] Danielle S. Bassett and Olaf Sporns. Network neuroscience, 2017.

[BSG+09] Stefan Bruckner, Veronika Soltészová, M Eduard Gröller, Jirí Hladůvka, Katja Bühler, Jai Y Yu, and Barry J Dickson. BrainGazer–visual queries for neurobiology research. *IEEE transactions on visualization and computer graphics*, 15(6):1497–504, 2009.

[BSJ+14] Joachim Böttger, Ralph Schurade, Estrid Jakobsen, Alexander Schäfer, and Daniel Margulies. Connexel visualization: a software implementation of glyphs and edge-bundling for dense connectivity data using brainGL. *Frontiers in Neuroscience*, 8:15, 2014.

[CBC+15] Evan Calabrese, Alexandra Badea, Gary Cofer, Yi Qi, and G Allan Johnson. A Diffusion MRI Tractography Connectome of the Mouse Brain and Comparison with Neuronal Tracer Data. *Cerebral cortex (New York, N.Y. : 1991)*, 25(11):4628–4637, 2015.

[CDW+16] Yuze Chi, Guohao Dai, Yu Wang, Guangyu Sun, Guoliang Li, and Huazhong Yang. NXgraph: An efficient graph processing system on a single machine. In *2016 IEEE 32nd International Conference on Data Engineering, ICDE 2016*, 2016.

[CVO+19] Etienne Combrisson, Raphael Vallat, Christian O'Reilly, Mainak Jas, Annalisa Pascarella, Anne-lise Saive, Thomas Thiery, David Meunier, Dmitrii Altukhov, Tarek Lajnef, Perrine Ruby, Aymeric Guillot, and Karim Jerbi. Visbrain: A Multi-Purpose GPU-Accelerated Open-Source Suite for Multimodal Brain Data Visualization. *Frontiers in Neuroinformatics*, 13:14, 2019.

[CYA+16] Giorgio Conte, Allen Q. Ye, Kyle R. Almryde, Olusola Ajilore, Alex D. Leow, and Angus Graeme Forbes. Intrinsic geometry visualization for the interactive analysis of brain connectivity patterns. In *Visualization and Data Analysis*, 2016.

[CYF+15] Giorgio Conte, Allen Q. Ye, Angus G. Forbes, Olusola Ajilore, and Alex Leow. BRAINtrinsic: A virtual reality-compatible tool for exploring intrinsic topologies of the human brain connectome. In *Brain informatics and health*, pages 67–76. Springer, 2015.

[DNM+18] Sergio Duca, Andrea Nani, Jordi Manuello, Karina Tatu, Tommaso Costa, Franco Cauda, Enrico Premi, Sara Palermo, and Peter T Fox. Brain structural alterations are distributed following functional, anatomic and genetic connectivity. *Brain*, 141(11):3211–3232, 2018.

[dRKY+18] Michael de Ridder, Karsten Klein, Jean Yang, Pengyi Yang, Jim Lagopoulos, Ian Hickie, Max Bennett, and Jinman Kim. An Uncertainty Visual Analytics Framework for fMRI Functional Connectivity. *Neuroinformatics*, 2018.

[dRvdH13] Marcel A de Reus and Martijn P van den Heuvel. The parcellation-based connectome: Limitations and extensions. *NeuroImage*, 80:397–404, 2013.

[FF16] Ben D. Fulcher and Alex Fornito. A transcriptional signature of hub connectivity in the mouse connectome. *Proceedings of the National Academy of Sciences*, 113(5):1435–1440, 2016.

[FGSZ85] M Ferrari, I Giannini, G Sideri, and E Zanette. *Continuous Non Invasive Monitoring of Human Brain by Near Infrared Spectroscopy*, pages 873–882. Springer US, Boston, MA, 1985.

[FLN+15] David Feng, Christopher Lau, Lydia Ng, Yang Li, Leonard Kuan, Susan M. Sunkin, Chinh Dang, and Michael Hawrylycz. Exploration and visualization of connectivity in the adult mouse brain. *Methods*, 73:90–97, 2015.

[FP11] Leon French and Paul Pavlidis. Relationships between gene expression and brain wiring in the adult rodent brain. *PLoS Computational Biology*, 7(1), 2011.

[Fra17] Lisa Frauenstein. Dimensionality reduction for analysis and visualization of functional connectivity in the developing human brain. Master's thesis, Fakultät für Informatik, Medical University of Vienna, Spitalgasse 23, A-1090 Vienna, Austria, 2017.

[fru19] Brain* for fruiflybrain, https://fruitfly.tefor.net, 2019.

[FSG+18] Lisa Frauenstein, Nicolas Swoboda, Florian Ganglberger, Georg Langs, and Katja Bühler. A prototypic visual analytics framework for interactive exploration of functional connectivity development (h023). In *Forum of Neuroscience of the Federation of European Neuroscience Societies, FENS 2018*, 2018.

[FTP11] Leon French, Powell Patrick Cheng Tan, and Paul Pavlidis. Large-Scale Analysis of Gene Expression and Connectivity in the Rodent Brain: Insights through Data Integration. *Frontiers in neuroinformatics*, 5:12, 2011.

[GDL+11] Stephan Gerhard, Alessandro Daducci, Alia Lemkaddem, Reto Meuli, Jean-Philippe Thiran, and Patric Hagmann. The Connectome Viewer Toolkit: An Open Source Framework to Manage, Analyze, and Visualize Connectomes. *Frontiers in Neuroinformatics*, 5:3, 2011.

[GKHB19] Florian Ganglberger, Joanna Kaczanowska, Wulf Haubensak, and Katja Bühler. A data structure for real-time aggregation queries of big brain networks. *Neuroinformatics*, pages 1–19, 2019.

[GKP+17] Florain Ganglberger, Joanna Kaczanowska, Josef Penninger, Andreas Hess, Katja Buehler, and Wulf Haubensak. In silico exploration of functional networks underlying behavioral traits. In *12th Göttingen Meeting of the German Neuroscience Society*, 2017.

[GKP+18] Florian Ganglberger, Joanna Kaczanowska, Josef M. Penninger, Andreas Hess, Katja Bühler, and Wulf Haubensak. Predicting functional neuroanatomical maps from fusing brain networks with genetic information. *NeuroImage*, 2018.

[GPB+18a] Johannes Griessner, Manuel Pasieka, Vincent Böhm, Florian Grössl, Joanna Kaczanowska, Pinelopi Pliota, Dominic Kargl, Barbara Werner, Nadia Kaouane, Sandra Strobelt, Silke Kreitz, Andreas Hess, and Wulf Haubensak. Central amygdala circuit dynamics underlying the benzodiazepine anxiolytic effect, 2018.

[GPB+18b] Johannes Griessner, Pinelopi Pilota, V. Boehm, M. Pasieka, F. Groessl, J. Kaczanowska, F. Ganglberger, and Haubensak W. Multi-modal exploration of a thalamo-amygdala anxiety circuit. In *Forum of Neuroscience of the Federation of European Neuroscience Societies, FENS 2018*, 2018.

[GSB+15] Erhan Genç, Marieke Louise Schölvinck, Johanna Bergmann, Wolf Singer, and Axel Kohler. Functional connectivity patterns of visual cortex reflect its anatomical organization. *Cerebral Cortex*, page bhv175, 2015.

[GSF+18] Florian Ganglberger, Nicolas Swoboda, Lisa Frauenstein, Joanna Kaczanowska, Wulf Haubensak, and Katja Bühler. Iterative exploration of big brain network data. In *VCBM 18: Eurographics Workshop on Visual Computing for Biology and Medicine, Granada, Spain, September 20-21, 2018*, pages 77–87, 2018.

[GSF+19] Florian Ganglberger, Nicolas Swoboda, Lisa Frauenstein, Joanna Kaczanowska, Wulf Haubensak, and Katja Buehler. Braintrawler: A visual analytics framework for iterative exploration of heterogeneous big brain data. *Computers & Graphics*, 82:304 – 320, 2019.

[GTF+18] F. Ganglberger, E. Tosun, L. Frauenstein, N. Swoboda, J. Kaczanowska, W. Haubensak, and K. Bühler. Braintrawler: A web-based visual analytics framework for big brain network data in their spatial context (h025). In *Forum of Neuroscience of the Federation of European Neuroscience Societies, FENS 2018*, 2018.

[Gut19] Florence Gutekunst. Guided data cleansing of large connectivity matrices. Master's thesis, Institute of Computer Graphics and Algorithms, Vienna University of Technology, Favoritenstrasse 9-11/186, A-1040 Vienna, Austria, 2019.

[HKBS07] Christopher J Honey, Rolf Kötter, Michael Breakspear, and Olaf Sporns. Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *Proceedings of the National Academy of Sciences*, 104(24):10240–10245, 2007.

[HKC+10] Wulf Haubensak, Prabhat S Kunwar, Haijiang Cai, Stephane Ciocchi, Nicholas R Wall, Ravikumar Ponnusamy, Jonathan Biag, Hong-Wei Dong, Karl Deisseroth, Edward M Callaway, et al. Genetic dissection of an amygdala microcircuit that gates conditioned fear. *Nature*, 468(7321):270–276, 2010.

[HLGB+12] Michael J Hawrylycz, Ed S Lein, Angela L Guillozet-Bongaarts, Elaine H Shen, Lydia Ng, Jeremy A Miller, Louie N van de Lagemaat, K A Smith, and Amanda et al. Ebbert. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature*, 489(7416):391–399, 2012.

[HLM+15] Anjanette P. Harris, Ross J. Lennen, Ian Marshall, Maurits A. Jansen, Cyril R. Pernet, Nichola M. Brydges, Ian C. Duguid, and Megan C. Holmes. Imaging learned fear circuitry in awake mice using fMRI. *European Journal of Neuroscience*, 2015.

[HLP+13] Wook-Shin Han, Sangyeon Lee, Kyungyeol Park, Jeong-Hoon Lee, Min-Soo Kim, Jinha Kim, and Hwanjo Yu. TurboGraph: A Fast Parallel Graph Engine Handling Billion-scale Graphs in a Single PC. *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2013.

[hum19] Human brain project, https://www.humanbrainproject.eu, 2019.

[HvM+17] Sjoerd Huisman, Baldur van Lew, Ahmed Mahfouz, Nicola Pezzotti, Thomas Höllt, Lieke Michielsen, Anna Vilanova, Marcel J.T. Reinders, and Boudewijn P.F. Lelieveldt. BrainScope: interactive visual exploration of the spatial and temporal human brain transcriptome. *Nucleic Acids Research*, 2017.

[ICT+12] Andrei Irimia, Micah C Chambers, Carinna M Torgerson, Maria Filippou, David A Hovda, Jeffry R Alger, Guido Gerig, Arthur W Toga, Paul M Vespa, Ron Kikinis, and John D Van Horn. Patient-tailored connectomics visualization for the assessment of white matter atrophy in traumatic brain injury. *Frontiers in neurology*, 3:10, 2012.

[JFD14] Shuiwang Ji, Ahmed Fakhry, and Houtao Deng. Integrative analysis of the connectivity and gene expression atlases in the mouse brain. *NeuroImage*, 84:245–253, 2014.

[JMR18] Chengtao Ji, Natasha M Maurits, and Jos B T M Roerdink. Data-driven visualization of multichannel EEG coherence networks based on community structure analysis. *Applied Network Science*, 3(1):41, 2018.

[KBG12] Aapo Kyrola, Guy Blelloch, and Carlos Guestrin. GraphChi: Large-Scale Graph Computation on Just a PC Disk-based Graph Computation. *Proceedings of the 10th USENIX conference on Operating Systems Design and Implementation*, 2012.

[KGG+19] Joanna Kaczanowska, Florian Ganglberger, Bence Galik, Andreas Hess, Yoshan Moodley, Katja Bühler, and Wulf Haubensak. Molecular archaeology of the human brain. *bioRxiv*, 2019.

[KGHB19] Joanna Kaczanowska, Florain Ganglberger, Wulf Haubensak, and Katja Buehler. Braintrawler: A web-based framework for iterative exploration of big brain network data. In *13th Göttingen Meeting of the German Neuroscience Society*, 2019.

[KGP+16] Joanna Kaczanowska, Florain Ganglberger, Josef Penninger, Andrea Hess, Katja Buehler, and Wulf Haubensak. In silico exploration of functional networks underlying behavioral traits. In *Forum of Neuroscience of the Federation of European Neuroscience Societies, FENS 2016*, 2016.

[Kin14] Merel Kindt. A behavioural neuroscience perspective on the aetiology and treatment of anxiety disorders. *Behaviour Research and Therapy*, 2014.

[KLVV13] S.-J. Kim, M Lewis, and J Veenstra-VanderWeele. Chapter 38 - The Developmental Neurobiology of Repetitive Behavior. In John L R Rubenstein and Pasko Rakic, editors, *Neural Circuit Development and Function in the Brain*, pages 761–782. Academic Press, Oxford, 2013.

[Koc04] Christof Koch. *The Quest for Consciousness: A Neurobiological Approach*. Roberts & Company Publishers, 2004.

[KZA+18] Johnson J G Keiriz, Liang Zhan, Olusola Ajilore, Alex D Leow, and Angus G Forbes. NeuroCave: A web-based immersive visualization

platform for exploring connectome datasets. *Network Neuroscience*, 2(3):344–361, 2018.

[lar19] Brain* for larvalbrain, http://www.larvalbrain.org, 2019.

[Laz10] Mariana Lazar. Mapping brain anatomical connectivity using white matter tractography. *NMR in biomedicine*, 23(7):821–835, 2010.

[LDTS14] Roan A LaPlante, Linda Douw, Wei Tang, and Steven M Stufflebeam. The Connectome Visualization Utility: Software for Visualization of Human Brain Networks. *PLoS ONE*, 9(12):e113838, 2014.

[LGF+12] Kaiming Li, Lei Guo, Carlos Faraco, Dajiang Zhu, Hanbo Chen, Yixuan Yuan, Jinglei Lv, Fan Deng, Xi Jiang, Tuo Zhang, Xintao Hu, Degang Zhang, L. Stephen Miller, and Tianming Liu. Visual analytics of brain networks. *NeuroImage*, 2012.

[LHA+07] Ed S Lein, Michael J Hawrylycz, Nancy Ao, Mikael Ayres, Amy Bensinger, Amy Bernard, Andrew F Boe, Mark S Boguski, Kevin S Brockway, Emi J Byrnes, et al. Genome-wide atlas of gene expression in the adult mouse brain. *Nature*, 445(7124):168–176, 2007.

[LM09] Stephen D Larson and Maryann E Martone. Ontologies for Neuroscience: What are they and What are they Good for? *Frontiers in neuroscience*, 3(1):60–67, 2009.

[LNT+08] Christopher Lau, Lydia Ng, Carol Thompson, Sayan Pathak, Leonard Kuan, Allan Jones, Mike Hawrylycz, P Carninci, T Kasukawa, S Katayama, and AW Toga. Exploration and visualization of gene expression with neuroanatomy in the adult mouse brain. *BMC Bioinformatics*, 9(1):153, 2008.

[LTW+11] Ching Yao Lin, Kuen Long Tsai, Sheng Chuan Wang, Chang Huain Hsieh, Hsiu Ming Chang, and Ann Shyn Chiang. The neuron navigator: Exploring the information pathway through the neural maze. In *IEEE Pacific Visualization Symposium 2011, PacificVis 2011 - Proceedings*, 2011.

[LY15] Feng-Mei Lu and Zhen Yuan. PET/SPECT molecular imaging in clinical neuroscience: recent advances in the investigation of CNS diseases. *Quantitative imaging in medicine and surgery*, 5(3):433–447, 2015.

[MBB+16] Sugeerth Murugesan, Kristofer Bouchard, Jesse A Brown, Bernd Hamann, William W Seeley, Andrew Trujillo, and Gunther H Weber. Brain Modulyzer : Interactive Visual Analysis of Functional Brain Connectivity. *IEEE/ACM transactions on computational biology and bioinformatics*, pages 1–14, 2016.

[MBWG13] Daniel S. Margulies, Joachim Böttger, Aimi Watanabe, and Krzysztof J. Gorgolewski. Visualizing the human connectome. *NeuroImage*, 80:445–461, 2013.

[MFH+05] Marcello Massimini, Fabio Ferrarelli, Reto Huber, Steve K Esser, Harpreet Singh, and Giulio Tononi. Breakdown of Cortical Effective Connectivity During Sleep. *Science*, 309(5744):2228–2232, 2005.

[MHLR16] Ahmed Mahfouz, Sjoerd M. H. Huisman, Boudewijn P. F. Lelieveldt, and Marcel J. T. Reinders. Brain transcriptome atlases: a computational perspective. *Brain Structure and Function*, pages 1–24, 2016.

[MKF+15] Chihua Ma, Robert V Kenyon, Angus G Forbes, Tanya Berger-Wolf, Bernard J Slater, and Daniel A Llano. Visualizing Dynamic Brain Networks Using an Animated Dual-Representation. In E Bertini, J Kennedy, and E Puppo, editors, *Eurographics Conference on Visualization (EuroVis) - Short Papers*. The Eurographics Association, 2015.

[MN08] S. S. Moy and J. J. Nadler. Advances in behavioral genetics: Mouse models of autism, 2008.

[mPlDI+16] Mu ming Poo, Jiu lin Du, Nancy Y. Ip, Zhi Qi Xiong, Bo Xu, and Tieniu Tan. China Brain Project: Basic Neuroscience, Brain Diseases, and Brain-Inspired Computing. *Neuron*, 2016.

[MvdGvdM+15] Ahmed Mahfouz, Martijn van de Giessen, Laurens van der Maaten, Sjoerd Huisman, Marcel Reinders, Michael J Hawrylycz, and Boudewijn P F Lelieveldt. Visualizing the spatial gene expression organization in the brain through non-linear similarity embeddings. *Methods*, 73:79–89, 2015.

[NRR+10] Laura J J Nicolaï, Ariane Ramaekers, Tim Raemaekers, Andrzej Drozdzecki, Alex S Mauss, Jiekun Yan, Matthias Landgraf, Wim Annaert, and Bassem A Hassan. Genetically encoded dendritic marker sheds light on neuronal connectivity in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America*, 107(47):20553–20558, 2010.

[NSvA+13] C Nowke, M Schmidt, S J van Albada, J M Eppler, Rembrandt Bakker, M Diesrnann, Bernd Hentschel, and T Kuhlen. VisNEST - Interactive analysis of neural activity data. *Biological Data Visualization (BioVis), 2013 IEEE Symposium on*, pages 65–72, 2013.

[OHN+14] Seung Wook Oh, Julie A Harris, Lydia Ng, Brent Winslow, Nicholas Cain, Stefan Mihalas, Quanxin Wang, Chris Lau, Leonard Kuan, Alex M Henry, et al. A mesoscale connectome of the mouse brain. *Nature*, 508(7495):207–214, 2014.

[PLK+15] Russell A. Poldrack, Timothy O Laumann, Oluwasanmi Koyejo, Brenda Gregory, Ashleigh Hover, and Mei-Yen et al. Chen. Long-term neural and physiological phenotyping of a single human. *Nature Communications*, 6:8885, 2015.

[Pol08] Russell A Poldrack. The role of fMRI in Cognitive Neuroscience: where do we stand? *Current Opinion in Neurobiology*, 18(2):223–227, 2008.

[RA15] Jonas Richiardi and Andre Altmann. Correlated gene expression supports synchronous activity in brain networks. *Science*, 348(6240):11–14, 2015.

[Rip19] Gwendolyn Rippberger. Data-driven anatomical layouting of brain network graphs, 2019.

[RLF15] Andre Santos Ribeiro, Luis Miguel Lacerda, and Hugo Alexandre Ferreira. Multimodal Imaging Brain Connectivity Analysis (MIBCA) toolbox. *PeerJ*, 3:e1078, 2015.

[RMZ13] Amitabha Roy, Ivo Mihailovic, and Willy Zwaenepoel. X-Stream: Edge-centric graph processing using streaming partitions. *Proceedings of the ACM Symposium on Operating Systems Principles*, 2013.

[RN13] Scott J. Russo and Eric J. Nestler. The brain reward circuitry in mood disorders. *Nature Reviews Neuroscience*, 2013.

[RS10] Mikail Rubinov and Olaf Sporns. Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*, 52(3):1059–1069, 2010.

[RYWB15] Mikail Rubinov, Rolf J. F. Ypma, Charles Watson, and Edward T. Bullmore. Wiring cost and topological participation of the mouse brain connectome. *Proceedings of the National Academy of Sciences*, 112(32):201420315, 2015.

[SCH+18] Johanna Senk, Corto Carde, Espen Hagen, Torsten W Kuhlen, Markus Diesmann, and Benjamin Weyers. VIOLA—A Multi-Purpose and Web-Based Visualization Tool for Neuronal-Network Simulation Output. *Frontiers in Neuroinformatics*, 12:75, 2018.

[SCHT09] Stephan Saalfeld, Albert Cardona, Volker Hartenstein, and Pavel Tomančák. CATMAID: Collaborative annotation toolkit for massive amounts of image data. *Bioinformatics*, 2009.

[SdS12] D L Schomer and Fernando da Silva. *Niedermeyer's electroencephalography: Basic principles, clinical applications, and related fields: Sixth edition.* 2012.

[SE12] Oliver Schmitt and Peter Eipert. neuroVIISAS: Approaching Multiscale Simulation of the Rat Connectome. *Neuroinformatics*, 10(3):243–267, 2012.

[SGHK03] Christoph Stosiek, Olga Garaschuk, Knut Holthoff, and Arthur Konnerth. In vivo two-photon calcium imaging of neuronal networks. *Proceedings of the National Academy of Sciences of the United States of America*, 100(12):7319–7324, 2003.

[SH09] Massimo Scanziani and Michael Häusser. Electrophysiology in the age of light. *Nature*, 461:930, 2009.

[SKI+17] Aslihan Selimbeyoglu, Christina K. Kim, Masatoshi Inoue, Soo Yeun Lee, Alice S. O. Hong, Isaac Kauvar, Charu Ramakrishnan, Lief E. Fenno, Thomas J. Davidson, Matthew Wright, and Karl Deisseroth. Modulation of prefrontal cortex excitation/inhibition balance rescues social behavior in cntnap2-deficient mice. *Science Translational Medicine*, 9(401), 2017.

[SKR+15] Tarek Sherif, Nicolas Kassis, Marc-Étienne Rousseau, Reza Adalat, and Alan C Evans. BrainBrowser: distributed, web-based neurological data visualization. *Frontiers in neuroinformatics*, 8:89, 2015.

[SMB+17] N Swoboda, J Moosburner, S Bruckner, J Y Yu, B J Dickson, and K Bühler. Visualization and Quantification for Interactive Analysis of Neural Connectivity in Drosophila. *Computer Graphics Forum*, 36(1):160–171, 2017.

[SMGL+16] Casey M Schneider-Mizell, Stephan Gerhard, Mark Longair, Tom Kazimiers, Feng Li, Maarten F Zwart, and Andrew et al. Champion. Quantitative neuroanatomy for connectomics in Drosophila. *eLife*, 5, 2016.

[SMT13] Niklas Smedemark-Margulies and Josef G Trapani. Tools, methods, and applications for optophysiology in neuroscience. *Frontiers in molecular neuroscience*, 6:18, 2013.

[Sor13] Johannes Sorger. neuroMap - Interactive Graph-Visualization of the Fruit Fly ' s Neural Circuit. In *2013 IEEE Symposium on Biological Data Visualization (BioVis)*, 2013.

[Spo13] Olaf Sporns. Structure and function of complex brain networks. *Dialogues in clinical neuroscience*, 15(3):247–262, 2013.

[vdEHBvW16] S van den Elzen, D Holten, J Blaas, and J J van Wijk. Reducing Snapshots to Points: A Visual Analytics Approach to Dynamic Network Exploration. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):1–10, 2016.

[VESB+13] David C Van Essen, Stephen M Smith, Deanna M Barch, Timothy EJ Behrens, Essa Yacoub, Kamil Ugurbil, WU-Minn HCP Consortium, et al. The wu-minn human connectome project: an overview. *Neuroimage*, 80:62–79, 2013.

[WVRG+16] Kirstie J Whitaker, Petra E Vértes, Rafael Romero-Garcia, František Váša, Michael Moutoussis, Gita Prabhu, Nikolaus Weiskopf, Martina F Callaghan, Konrad Wagstyl, Timothy Rittman, Roger Tait, Cinly Ooi, John Suckling, Becky Inkster, Peter Fonagy, Raymond J Dolan, Peter B Jones, Ian M Goodyer, the NSPN Consortium, and Edward T Bullmore. Adolescence is associated with genomically patterned consolidation of the hubs of the human brain connectome. *Proceedings of the National Academy of Sciences*, 113(32):9105–9110, 2016.

[XWH13] Ming-Rui Xia, Jin-Hui Wang, and Yong He. BrainNet Viewer: a network visualization tool for human brain connectomics. *PloS one*, 8(7), 2013.

[YR09] Yaling Yang and Adrian Raine. Prefrontal structural and functional brain imaging findings in antisocial, violent, and psychopathic individuals: a meta-analysis. *Psychiatry research*, 174(2):81–88, 2009.

[YW04] Larry J Young and Zuoxin Wang. The neurobiology of pair bonding. *Nature neuroscience*, 7(10):1048–1054, 2004.

[zeb19] Brain* for zebrafishbrain, https://zebrafish.tefor.net/, 2019.

[ZFB10] Andrew Zalesky, Alex Fornito, and Edward T. Bullmore. Network-based statistic: Identifying differences in brain networks. *NeuroImage*, 53(4):1197–1207, 2010.

[ZLP+18] Zhihao Zheng, J Scott Lauritzen, Eric Perlman, Camenzind G Robinson, Matthew Nichols, Daniel Milkie, Omar Torrens, John Price, Corey B Fisher, Nadiya Sharifi, Steven A Calle-Schuler, Lucia Kmecova, Iqbal J Ali, Bill Karsh, Eric T Trautman, John A Bogovic, Philipp Hanslovsky, Gregory S X E Jefferis, Michael Kazhdan, Khaled Khairy, Stephan Saalfeld, Richard D Fetter, and Davi D Bock. A Complete Electron Microscopy Volume of the Brain of Adult Drosophila melanogaster. *Cell*, 174(3):730–743.e22, 2018.

[ZMB+15] Da Zheng, Disa Mhembere, Randal Burns, Joshua Vogelstein, Carey E Priebe, and Alexander S Szalay. FlashGraph: Processing Billion-node Graphs on an Array of Commodity SSDs. In *Proceedings of the 13th USENIX Conference on File and Storage Technologies*, FAST'15, pages 45–58, Berkeley, CA, USA, 2015. USENIX Association.

Curriculum Vitae

Europass Curriculum Vitae

Personal information

First name(s) / Surname(s)	**Florian Johann Ganglberger**
Nationality	Austrian
birth	1989

Work experience

Dates	Since February 2014
Occupation or position held	PhD Student / Researcher
Main activities and responsibilities	Biomedical Visualization and Image Processing | 40h per week
Employer	VRVis Research Center for Virtual Reality and Visualization
Type of business or sector	Research

Dates	From November 2012 to December 2013
Occupation or position held	Java Software Developer
Main activities and responsibilities	Metabolomic pathway visualization in Java | 20h per week
Employer	Biocrates Life Sciences AG
Type of business or sector	Bioinformatics / Research

Dates	From October 2012 to January 2014
Occupation or position held	PhD Student
Main activities and responsibilities	Computerized recognition of anatomical objects: A feasibility study | 20h per week Student representative of the PhD programme “Image Guided Diagnosis and Therapy”
Employer	4D Visualization Lab, University Clinic of Oto-, Rhino-Laryngology
Type of business or sector	Medical Image Analysis / Research

Dates	From May 2012 to September 2012
Occupation or position held	Technical Assistant
Main activities and responsibilities	Statistical Analysis and Visualisation in R and Java | 30h per week Common language was English due to an international team
Employer	Centre for Molecular Medicine of the Austrian Academy of Sciences
Type of business or sector	Bioinformatics / Research

For more information on Europass go to http://europass.cedefop.europa.eu
© European Union, 2004-2010 24082010

Dates	From January 2012 to May 2012
Occupation or position held	Diploma Student
Main activities and responsibilities	Non-linear Mapping of Drosophila Populations based on Neuronal Structure | 40h per week
Employer	Computational Image Analysis and Radiology Lab of the Medical University of Vienna
Type of business or sector	Medical Image Analysis / Research
Dates	From August 2010 to January 2012
Occupation or position held	Technical Assistant
Main activities and responsibilities	Statistical Analysis and Visualisation in R and Java | 15h per week Common language was English due to an international team
Employer	Centre for Molecular Medicine of the Austrian Academy of Sciences
Type of business or sector	Bioinformatics / Research
Dates	September 2011
Occupation or position held	Internship
Main activities and responsibilities	Structure-based local search of neurons in Drosophila brain | 30h per week
Employer	Computational Image Analysis and Radiology Lab of the Medical University of Vienna
Type of business or sector	Medical Image Analysis / Research
Dates	From April 2010 to April 2011
Occupation or position held	Web-Developer
Main activities and responsibilities	Development of Content-Management-System based Websites with PHP and CSS | 10h per week
Employer	elecks.com
Type of business or sector	Advertisement
Dates	From April 2009 to April 2010
Occupation or position held	Event-Photograph
Main activities and responsibilities	Photography of local events | 5-10h on weekends
Employer	Einfachwerben
Type of business or sector	Advertisement
Dates	August 2009
Occupation or position held	Internship
Main activities and responsibilities	Visualisation of Vehicle-Tracking data with OpenStreet Maps and Java | 40h per week
Employer	Federal-State Government Upper-Austria, Department of Overall Traffic-Planning
Type of business or sector	Traffic-Planning
Dates	From July 2008 to December 2008
Occupation or position held	Soldier
Main activities and responsibilities	Vehicle Operator
Employer	Department of Defence, Austria
Type of business or sector	Military Service
Dates	From July 2007 to August 2007

For more information on Europass go to http://europass.cedefop.europa.eu
© European Union, 2004-2010 24082010

Occupation or position held	Internship
Main activities and responsibilities	3D Surface and Image Analysis in C#.net/C++ | 40h per week
Employer	HLC Coating GmbH
Type of business or sector	Surface-Coating for Automobile-Industry
Dates	From July 2006 to August 2006
Occupation or position held	Internship
Main activities and responsibilities	Macro Development for ARCHline XP in C++ | 40h per week
Employer	IT-Concept Software GmbH
Type of business or sector	Software for Architects
Dates	July 2005
Occupation or position held	Internship
Main activities and responsibilities	Development of a Website, System Administration | 40h per week
Employer	Office for Planning and Civil Engineering Ganglberger
Type of business or sector	Civil Engineering
Dates	July 2004
Occupation or position held	Labourer
Main activities and responsibilities	Working on a Building Site | 40h per week
Employer	Krückl GmbH and CoKG
Type of business or sector	Civil Engineering

Education and training

Dates	Since February 2014
Title of qualification awarded	Doktor der technischen Wissenschaften (PhD in Medical Informatics)
Principal subjects/occupational skills covered	- Biomedical Visualization - Biomedical Image Processing - Large scale data analysis
Name and type of organisation providing education and training	Technical University of Vienna
Dates	October 2012 – January 2014
Title of qualification awarded	Discontinued
Principal subjects/occupational skills covered	- PhD Program Image Guided Diagnosis and Therapy - Intraoperative 3D navigation in surgery - Medical Image Processing - Augmented Reality applications for stereoscopic images
Name and type of organisation providing education and training	Medical University of Innsbruck
Dates	From March 2011 to September 2012
Title of qualification awarded	Master of Science in Medical Informatics (grade average 1.0)
Principal subjects/occupational skills covered	- Pattern Recognition - Medical Image Processing - Advanced knowledge of statistics in medical and biological studies
Name and type of organisation providing education and training	Technical University of Vienna

For more information on Europass go to http://europass.cedefop.europa.eu
© European Union, 2004-2010 24082010

Dates	From March 2009 to March 2011
Title of qualification awarded	Bachelor of Science in Medical Informatics (grade average 1.5)
Principal subjects/occupational skills covered	- The computer-assisted processing of medical data - Simulation of biological processes and the use of imaging processes (e.g. computer tomography) in diagnosis or operational interventions are examples of the use of medical IT - Basics in medical principles such as biochemistry, anatomy, pathology and medical methodology
Name and type of organisation providing education and training	Technical University of Vienna

Dates	From September 2003 to June 2008
Title of qualification awarded	Matura (Austrian School Leaving Certificate)
Principal subjects/occupational skills covered	- General IT Knowledge - Commercial Data Processing - Organization and Business Informatics - Software Engineering
Name and type of organisation providing education and training	HTL Perg (Technical Highschool)

Personal skills and competences

Mother tongue(s) **German**

Other language(s)

Self-assessment	Understanding				Speaking				Writing	
European level (*)	Listening		Reading		Spoken interaction		Spoken production			
English	(C1)	Independent user	(C1)	Independent user	(C1)	Independent user	(C1)	Independent user	(C1)	Independent user

(*) Common European Framework of Reference for Languages

Social skills and competences	- good ability to adapt to a multicultural team - good communication skills gained through my experience as event photographer
Technical skills and competences	- Statistical analysis - Explorative data analysisand visualization - Neuroinformatics/Neurobiology - Pattern recognition and Image processing - Machine learning - Parallel computing
Computer skills and competences	- excellent skills in programming languages like C, C++, C#, .net Framework, Java, GO - excellent skills in R and Matlab - good skills in Web-development (HTML, CSS, PHP, AJAX, Javascript, reactJS, WebGL) - good skills in Databasesystems (SQL,PL/SQL)
Driving licence	Category B (Austria/EU)

For more information on Europass go to http://europass.cedefop.europa.eu
© European Union, 2004-2010 24082010