DISCLAIMER: More detailed information on the results and/or performance obtained and their use is available in the Subproject's subsequent Deliverable (D1.4.4) and/or Periodic Report.

Project Number:	604102	Project Title:	Human Brain Project

Document Title:	Strategic Mouse Data for the HBP Mouse Brain Atlas: Package one (data)
Document Filename:	SP1 D1.4.3 FINAL.docx
Deliverable Number:	D1.4.3
Deliverable Type:	Data
Work Package(s):	WPs 1.1, 1.2, 1.3, 1.4 (WPs involved in writing this document)
Dissemination Level:	PU = Public
Planned Delivery Date:	M18 / 31 Mar 2015
Actual Delivery Date:	M19 / 01 Sep 2015

Authors:	UPM (P59): Pilar F. ROMERO, Javier DEFELIPE UEDIN (P71): Elaine MARSHALL, Seth GRANT
Compiling Editors:	Pilar F. ROMERO, UPM (P59); Elaine MARSHALL, UEDIN (P71)
Contributors:	LENS (P34): Francesco PAVONE, Ludovico SILVESTRI IEM HAS (P28): Tamás FREUND, Szabolcs KALI UAM (P57): Francisco CLASCA, UPM (P59): Ángel MERCHÁN UZH (P63): Bruno WEBER, UEDIN (P71): Marcia ROY, David SERRATT, Douglas ARMSTRONG WMC (P79): Yun WANG, UOXF (P8): Chris PONTING, Grant BELGARD
STO Review:	UHEI (P45): Björn KINDLER, Sabine SCHNEIDER, Martina SCHMALHOLZ
Editorial Review:	EPFL (P1): Guy WILLIS, Lauren ORWIN

Abstract:	This report provides a detailed account of the strategic mouse data for the HBP Mouse Brain Atlas. Within this report, we outline the data that will be deposited in the HBP Mouse Brain Atlas. This data includes the first draft transcriptomes types for major neocortical neurons, the first draft generic synapse proteome, first whole brain scans of the distribution of selected proteins, first whole brain scans of axonal projections from selected brain regions, first whole brain scan of synapses with quantification in selected brain regions, first high resolution whole brain scans of the vasculature, first numbers and distributions of neurons and glia, first cell distribution maps of the whole mouse brain, and first morphologies of neurons and glia in hippocampus and neocortex. A short report will describe each level of data.
Keywords:	Single cell transcriptomics, synapse proteome, axonal projections, synaptic maps, ultrastructural data, brain vasculature, cell distributions, cell morphologies.

Table of Contents

List of Figures and Tables

1. Introduction

This report includes the first SP1 datasets that will be deposited in the HBP Mouse Brain Atlas. D1.4.3 describes each level of data and provides quantitative indicators of the completeness of the data, compared to the targeted dataset and to a projected full dataset to be generated by the research community. This data includes the first:

- Draft transcriptomes types for major neocortical neurons
- Draft generic synapse proteome
- Whole brain scans of the distribution of selected proteins
- Whole brain scans of axonal projections from selected brain regions
- Whole brain scan of synapses with quantification in selected brain regions
- High-resolution, whole-brain scans of the vasculature
- Numbers and distributions of neurons and glia
- Cell distribution maps of the whole mouse brain and first morphologies of neurons and glia in hippocampus and neocortex.

Each type of data is described below in more detail.

1.1 Brief overview of data and its scientific significance by the SP Leader

During Months 1—18, the Molecular Work Packages (WP1.1 and 1.3) established robust and reproducible protocols that delivered the first draft neocortical transcriptome and the first draft of the trans-synaptic proteome across the mouse brain. A pipeline for data registration and integration into the Neuroinformatics Platform has also been established.

The data integration activities of the four Tasks that started in April 2014 (WP1.1) are not part of this Deliverable. For these Tasks, data delivery is from Month 24 onwards. The work progress of the four new tasks and their integration in SP1 is described in detail in the Month 18 periodic report and the corresponding Milestones section in EMDESK.

In the same period, the Cellular Work Package (WP1.2) has made significant progress in determining maps of the brain vasculature, morphologies of key mouse neurons, principles of long-range projections, and the distribution of neurons and synapses. The data and tools that have been generated by the Cellular Work Package represent crucial steps towards the construction of the first mouse brain models.

1.2 Description of how the data link to the Neuroinformatics Platform

All the groups in SP1 have used a data registration process that links data generated within SP1, plus data curated from the literature, into the central HBP Data Register. This means that both the raw datasets and annotated datasets are logged with key terms and appropriate metadata. At all times, data provenance is recorded, so that a search on an annotated data will reveal both the annotator and the original data contributor. This is an important step in engaging the wider HBP community through making our datasets searchable and available.

2. Strategic Mouse Data for the HBP Mouse Brain Atlas

2.1 First Draft Transcriptome Types for Major Neocortical Cells

Responsible: Task T1.1.2 (UOXF)

2.1.1 Data Description

We have met our Milestone MS3 (First draft transcriptome of major neocortical neurons; first draft of synapse proteome) through groundbreaking work by SP1 partner Sten Linnarsson. Our transcriptomics Milestones would not have been met without considerable cooperation within the Subproject in developing assays and sharing methods and data.

These data comprise shotgun-captured single-cell transcriptomes from various cell types from *post-mortem* primary somatosensory cortex of outbred CD1 mice, of age postnatal day 21-31, both male and female. The format of the data provided to the Neuroinformatics Platform (SP5) is a cell type-by-gene matrix giving the average number of transcripts from each gene in cells of that type. This work used the Fluidigm C1 system and the STRT method, which pairs unique molecular identifiers with sequencing of transcript tags.

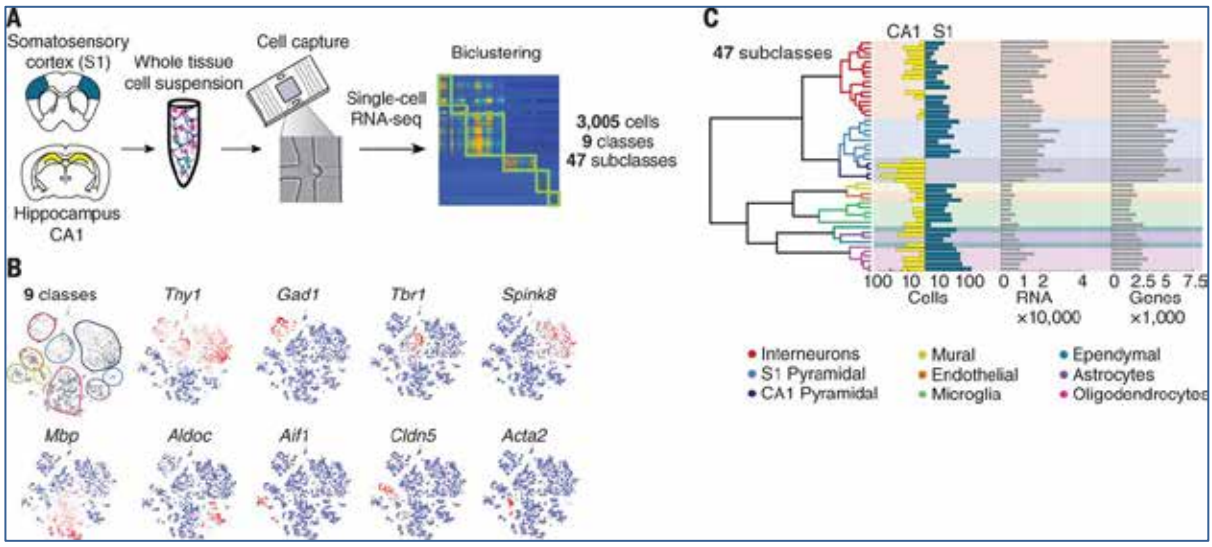

Figure 1: Molecular census of somatosensory S1 cortex and hippocampus CA1 by unbiased sampling and single-cell RNA-seq

Figure 1 shows the workflow for obtaining and analyzing single-cell RNA-sequences from juvenile mouse cortical cells, from dissection to single-cell RNA-sequencing and biclustering (Fig. 1, A) Visualization of nine major classes of cells using t-distributed stochastic neighbor embedding (tSNE). Each dot is a single cell, and cells are laid out to show similarities. Coloured contours correspond to the nine clusters shown in Fig. 1 B. Expression of known markers is shown in Fig. 1 C, using the same colour coding (blue, no expression; white, 1% quantile; red, 99% quantile), with hierarchical clustering analysis of 47 subclasses. Bar plots show the number of captured cells in CA1 and S1, the number of detected polyA+ RNA molecules per cell, and the total number of genes detected per cell. Fig.1 is from Zeisel *et al.* (2015).

These data will be deposited into the Mouse Brain Atlas within the Neuroinformatics Platform (SP5) and used in the Brain Simulation Platform (SP6) to initialise cellular models with non-uniform priors on the distributions of molecules of key physiological interest across various cell types. The relative levels among cell types should be especially informative here.

2.1.2 Quantitative indicators of data completeness compared to the targeted data set and a projected full data set to be generated by the research community:

SP1 expected to deliver six neuronal type transcriptomes from neocortex in M18.

Through single-cell sequencing of 3,005 shotgun-captured mouse cells from S1 of cortex and CA1 of hippocampus, eight pyramidal cortical types with specific laminar patterning and fourteen interneuron types found at least partially in S1 were identified, including a previously unknown neuronal type in cortical layer 1. This exceeds the six cortical neuronal transcriptomes that were planned for M18. After initial clustering of cells based on their transcriptomes, markers of these cell types were identified. *In situ* hybridizations of these marker genes were then used to assess their anatomical distributions and to match them to known cell types of known morphology. Multiple glial types were found as well, for a total of 47 delineated types. We detected approximately 20,000 transcripts expressed from approximately 5,000 genes in the neuronal cell types and approximately 5,000-10,000 transcripts from 2,500 genes in the glial cell types. Due to imperfect efficiency, this is a lower bound. It is difficult to determine this efficiency from current data, as it varies considerably from gene to gene. The distributions of these types are not representative of the expected relative representations of cells *in vivo*.

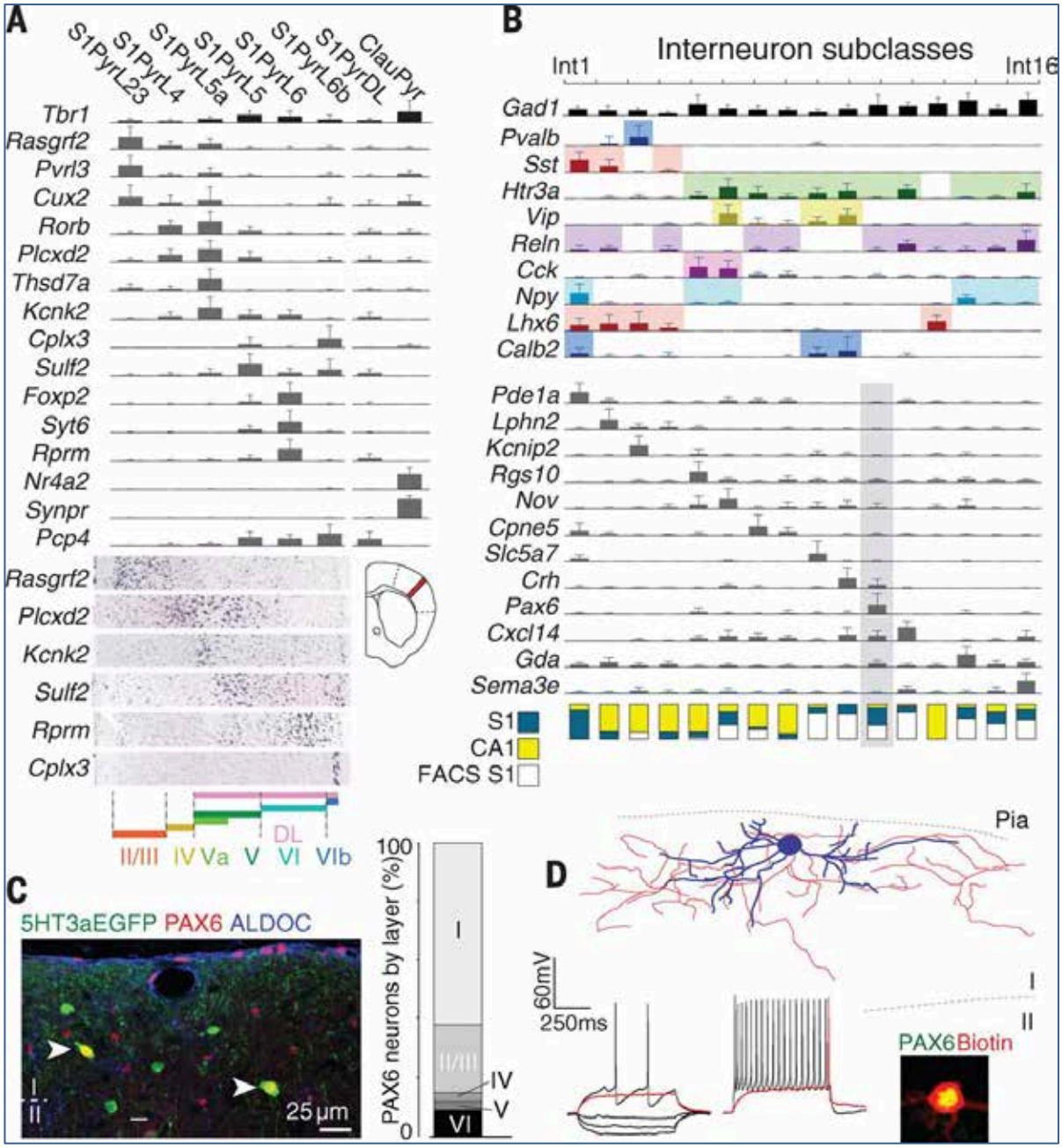

Figure 2: Neuron subclasses in the somatosensory cortex

In Figure 2, subclasses of pyramidal neurons in the somatosensory cortex (S1) identified by clustering. Bar plots show mean expression of selected known and novel markers (error bars show standard deviations). Layer-specific expression shown by in situ hybridization (Allen Brain Atlas). S1PyrL23, layer II-III; S1PyrL4, layer IV; S1PyrL5a, layer Va; S1PyrL5, layer V; S1PyrL6, layer VI; S1PyrL6b, layer VIb; S1PyrDL, deep layers; ClauPyr, claustrum. (**B**) Identification of interneuron subclasses. Bar plots show selected known and novel markers. Fraction of S1/CA1 cells is depicted at bottom: blue, S1; yellow, CA1; white, flow-sorted Htr3a$^+$ cells from S1. (**C**) Immunohistochemistry demonstrating the existence and localization of novel PAX6$^+$/5HT3aEGFP$^+$ interneurons, Int11. Bar plots show the layer distribution of these neurons. (**D**) Intrinsic electrophysiology and morphology of PAX6$^+$ interneurons in S1 layer I, identified by post hoc staining. Figure and legend are from Zeisel *et al.* (2015).

We anticipate that the larger research community will produce a wide variety of single cell and cell-type transcriptomes in the coming years, but it is unlikely that key variables such as strain and age will be aligned with other datasets being produced within HBP. The hippocampal portions of this dataset overlaps with Milestone MS4. MS4 will be rounded out

with additional single-cell transcriptomes from the remaining hippocampal regions, using the full-length Smart-seq2 method on shotgun-captured cells on the C1 from Chris Ponting's lab (Oxford). The Smart-seq2 method on the C1 will also be used to generate data from cerebellum in Chris Ponting's lab. We anticipate that additional non-strategic single-cell and single-cell-type transcriptome data will come from the Allen Institute in the USA and from the laboratories of Nathaniel Heintz (Rockefeller) and Myriam Heiman (MIT), all derived using the BAC TRAP method.

These data have been provided to the Neuroinformatics Platform (SP5), but we have not yet been provided with a link to access the data on the Neuroinformatics Platform. We are currently working with the Neuroinformatics Platform team to assemble a table for translation between these protein identifiers and the gene identifiers used in these data. We have been cooperating across Platforms by identifying a shortlist of proteins of interest with modellers in other Subprojects.

2.1.3 Data Provenance

Methods are described in great detail in "Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq" by Zeisel *et al*. Science 6 March 2015[1].

2.2 First Draft Generic Synapse Proteome

Responsible: Task T1.1.3 (UEDIN)

2.2.1 Data Description

During the period Month 13—18, significant progress has been achieved in T1.1.3.

Completion of Milestones: UEDIN has achieved MS3: First draft of the generic synapse mouse brain proteome across the whole brain of 2-month old male C57BL/6j mice. Briefly, UEDIN developed micro-scaled synaptosome enrichment techniques from as little as 20 mg (one single hippocampus) of mouse brain tissue. These techniques, which were optimised in Months 1-12 and described in D1.4.2, were applied to dissected mouse brain tissue from 7 integral regions of the mouse brain; striatum, hippocampus, hypothalamus, cerebellum, medial cortex, caudal cortex and frontal cortex from n=6 two month old adult mice and enabled the identification of ~700 proteins by LC-MS/MS including HBP priority proteins.

In total, 42 synaptic enrichments were analysed by LC-MS/MS and a total of 2,108 proteins have been identified in the young adult mouse brain synaptic proteome. Label-free quantitation of all identified proteins has identified differentially abundant proteins in all 7 integral regions: striatum (150), hippocampus (91), hypothalamus (233), cerebellum (238), frontal cortex (14), medial cortex (66) and caudal cortex (31). The MS data obtained for each brain region is in the form of Excel spreadsheets containing 2 tabs of information. The first tab contains the Mascot search results of the raw MS data for protein identification and the second tab contains the label-free quantitation of MS results using Progenesis software (Linear Dynamics). The results of the label-free relative quantitation across all brain regions examined for all n=6 mice is presented in an Excel spreadsheet and the data is organized under the following columns headings: (protein) Accession, Peptide count, Unique peptide count, Confidence score, ANOVA p-value), Maximum fold change (in protein abundance), Highest mean abundance (across all regions), Lowest mean abundance (across all regions), protein description, normalized abundance (MS spectral intensity). The entire dataset has also been searched for phosphorylation and acetylation modifications and will be ready for dissemination to HBP partners shortly.

Having obtained the first map of the synaptic proteome across the seven integral regions of the mouse brain, we are now examining the dataset for differences in proteome regulation, i.e., phosphorylation and acetylation. The methodology for the dissection of entire mouse brains by micropunch has been established, in collaboration with Javier de Felipe's laboratory. This second method will allow for high-resolution mass spectrometry-based proteomic profiles to be obtained for each $1mm^3$ volume throughout the entire mouse brain.

Our close collaboration with UPM has led to the successful generation of the first generic draft of the synapse proteome across 7 integral regions of the mouse brain. It has also led to the development of finer sampling techniques of various regions across the entire mouse brain, as outlined above. This will lead to the generation of a finer detailed atlas of the mouse brain synaptic proteome, comparable to the one generated by the Allen Brain Institute at the RNA-level.

All datasets obtained from all proteomic pilot studies performed since June 2014, as well as the first draft of the synaptic proteome across the mouse brain, have been provided to WP1.3 (Armstrong, UEDIN) and are in the process of being shared within the Neuroinformatics Platform (SP5).

We anticipate that additional proteome data may be generated by the research community. Strategic data will be identified by WP1.3 and provided to the

Neuroinformatics Platform. In collaboration with T1.1.2, WP1.3, and modellers in other Subprojects, we will continue to identify proteins of interest.

2.2.2 Data Provenance

All Excel spreadsheet data files that have been shared with the Neuroinformatics Platform include tabs and column sub-headings under both raw abundance and normalized abundance, in which the individual MS datafile is listed. Each sample analysed by the Kinetic Paramater Facility (KPF), Edinburgh, has a unique MS run file number generated and associated with it. Therefore, all data generated by UEDIN can be identified by its listed unique MS run file which is documented and stored in the Grant lab UEDIN as well as the KPF in Edinburgh.

2.3 First Brain Scans of Axonal Projections from Selected Brain Regions

Responsible: Task 1.2.5 (UAM)

2.3.1 Data Description

- Species: mice (*mus musculus*, male, C57BL6 strain, 60—90 days old)
- Brain region: Visual cortical areas and related thalamic nuclei (DLG, LP), and Somato-motor cortical areas and related thalamic nuclei (Po VPM)
- Cell type: Thalamocortical projection cells neurons and VTA projection neurons.
- Origin (e.g. *in vitro*, *post mortem*): Inmunohistochemically-enhanced (DAB-nickel) labelling with Pal-GFP (Kuramoto et al., Cerebral Cortex 19:2065-77, 2009[2]) following single-cell transfection in vivo.
- Data format: Tiff 2D drawings and Neurolucida 3D files (.DAT). Additional Tiff files of selected tissue sections form each experiment will also be provided.

In T1.2.5, three different methods for labelling in their entirety the axoendritic tree and the multi-ramified axon of Long-Range Projection Neurons (LRPNs) have been tested and compared extensively: a) juxtacellular BDA injections; b) low titration viral vector transfection; and c) electroporation of viral nucleic acids). T1.2.5 has opted for the last one, which is the most versatile and reliable.

Four methods for tracing and 3D-reconstructing have been examined and/or experimentally tested: a) Serial Two-Photon Scanning of fluorescently labelled LRPNs (Ragan et al. Nature Communications 2012) c; b) Laser-Sheet Scanning of fluorescent LRPNs cleared tissue blocks (review in Keller & Ahrens, Neuron 2015); c) Imaging serial microtomy of resin embedded fluorescent tissue (Gong et al., Neuroimage 2013[3]); and d) and serial freeze sectioning + inmunohistochemical amplification and conversion of the fluorescence into a photostable opaque labelling amenable for intensive high-magnification analysis. T1.2.5 has opted for the last one of these strategies.

The Neuroinformatics (SP5) and the Brain Simulation (SP6) Platforms are currently being given access to the delivered data set. The axonal and dendritic architecture and a series of tissue landmarks of the labelled cells is being digitally reconstructed from the serial sections. The specific file formats and ancillary information have been discussed and agreed with SP5 and SP6.

In T1.2.1, the tracing of single-cell axonal projections is scheduled to reach 100% and the data set that is expected to be delivered comprises 30 axonal projections traced by Month

30. In addition, the determination of the distribution of terminal axons branches and quantification of monosynaptic targets is scheduled to reach 100%, and the data set that is expected to be delivered should include 40 terminal axon branches and identified monosynaptic targets.

By Month 18, we had 2D-traced 11 individual LRPNs, and 3D-traced 2 of them. That is in line with the expected output of our Task at this juncture. It should be pointed out that that the main thrust of this Task effort in Month 1-18 was not focused on data production, but rather on testing and selecting the most adequate experimental tools – among the several available - as well as on defining the relevant data parameters and digital formats for establishing a consensus HBP protocol for generating and recording LRPN morphologies. It should be remembered that the complex LRPN neurons constitute a sizable proportion of brain cells and are the key to the integrated, small-world properties of brain networks, but have not been before systematically labelled or studied at the single cell level. T1.2.5 is currently very close to finishing the testing and definition of this consensus protocol.

It is expected that, once agreed, this protocol will become the HBP guide for the production, gathering and indexation of reliable, quantitative morphological data on LPRNs in all long range brain pathways, in a digital format that can be readily shared across HBP Platforms. Moreover, the quality of cell labelling and the quantitative precision in its 3D indexation to the digital mouse brain atlas are crucial for being relevant to the predictive circuit modelling strategy of HBP.

2.3.2 Data Provenance

Data generated have been captured on site at Clasca's laboratory.

2.4 First Brain Scan of Synapses with Quantification in Selected Brain Regions

Responsible: T1.2.6 (UPM)

2.4.1 Data Description

- Species: *mus musculus*, PSD95GFP strain
- Brain region: hippocampus and somatosensory cortex
- Origin: *post mortem*
- Data format: tiff image stacks and excel files

We are using PSD95GFP mice to study the density and distribution of PSD95 puncta in the hippocampus and somatosensory cortex in fixed brains, in collaboration with Grant's laboratory. Data have been delivered as tiff image stacks and excel files. The format of these data has been aligned with the Neuroinformatics Platform (SP5) requirements.

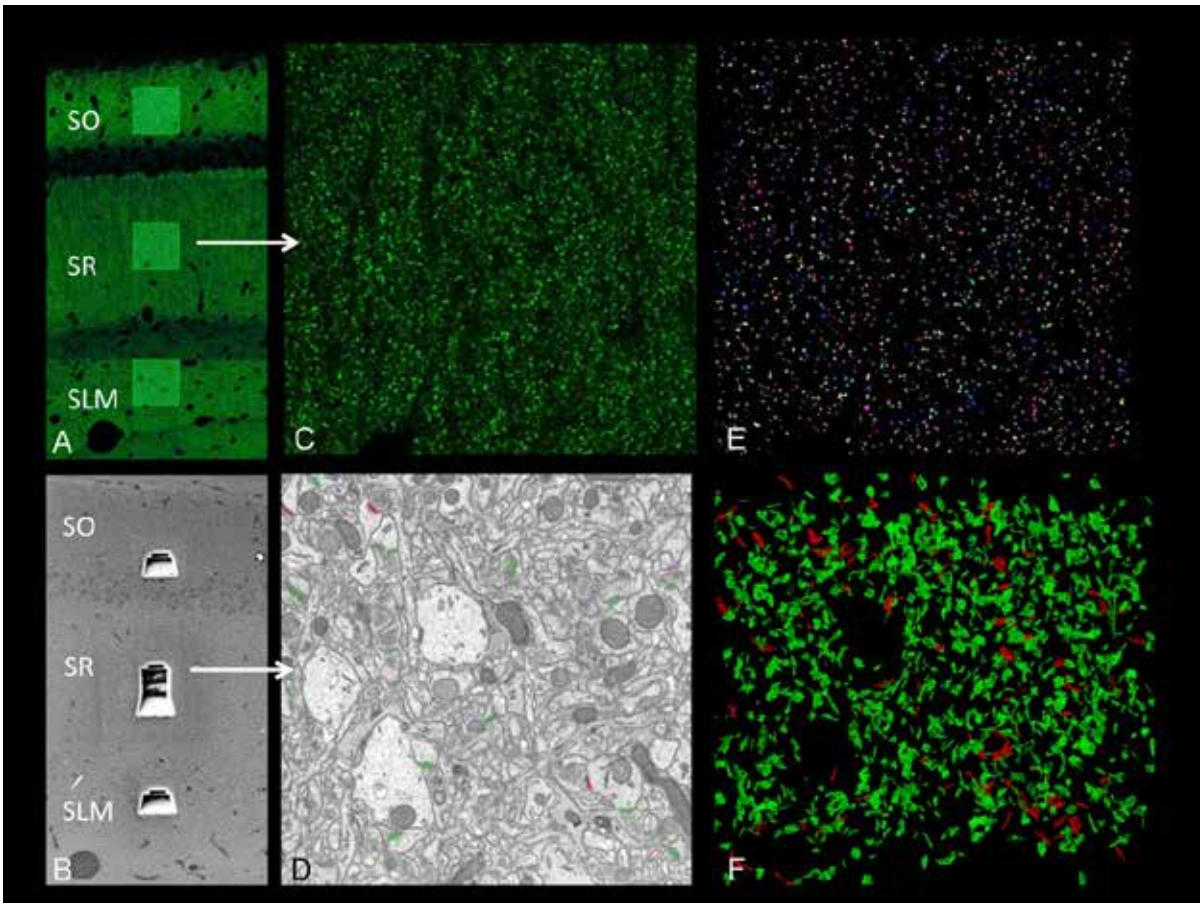

Figure 3: (A, B) Low-magnification confocal (A) and FIB-SEM (B) images of adjacent sections of the CA1 hipocampal formation of PSD95 EGFP mice.

This illustrates different regions where images were acquired (SO: stratum oriens; SR: stratum radiatum; SLM: stratum lacunosum moleculare). (C, D) Confocal (C) and FIB-SEM (D) image stack examples to illustrate PSD95 labeling (green in C) and the postsynaptic densities (PSD) of the neuropil (in D: green indicates excitatory identified synapses; red indicates inhibitory identified synapse). (E) Example of 3D segmented puncta from confocal image stack shown in C. Each color was randomly assigned to visualize individually detected PSD. (F) Example of 3D segmented PSD from FIB-SEM image stack shown in D.

In this Task, the development of 3D methods was scheduled to be completed by Month 18 and the generation of synaptic maps and ultrastructural data was scheduled to reach 100% by Month 30. The planned data set comprises 60 synaptic maps of individual identified neurons and two synaptic maps of mouse brain by Month 30.

By Month 18, the 3D methods development had been achieved as planned and 10% of the generation of synaptic maps and ultrastructural data had been accomplished - ahead of the planned schedule. Moreover, the data set generated already features 12 synaptic maps of individual identified neurons and two synaptic maps of mouse brain.

The strategic data being generated by this Task will feed the Neuroinformatics Platform (SP5) and, indirectly, the Brain Simulation (SP6), Neuromorphic Computing (SP9) and Neurorobotics (SP10) Platforms.

2.4.2 Data Provenance

The data generated are from Grant's & DeFelipe's laboratories

2.5 First High Resolution Brain Scans of the Vasculature

Responsible: T1.2.1 (UZH)

2.5.1 Data Description

- Species: *mus musculus*
- Brain region: vasculature (synchrotron)
- Origin: *post mortem*
- Data format: Image stacks, meshes

The generation of mouse vasculature maps by T1.2.1 is scheduled be completed with the delivery of six mouse vasculature maps by Month 30.

By Month 18, T1.2.1 had completed its reconstruction methods, and has achieved 35% of the generation of mouse vasculature maps, delivering six maps (see table 1 at the end of this chapter).

The full strategic data set likely to be generated in the long term in this Task will feed the Platforms.

The Neuroinformatics Platform (SP5) has no access to the delivered data as yet, but the process for achieving this is underway.

2.5.2 Data Provenance

The data generated have been captured on site at Weber's laboratory.

2.6 First Cell Distribution Maps of the Whole Mouse Brain

Responsible: T1.2.2 (LENS)

2.6.1 Data Description

- Species: *mus musculus*
- Brain region: cerebellum
- Cell type: Purkinje cells
- Origin: *post mortem*
- Data format: Image stacks (3D tiff), point cloud (vtk), csv

The format of the data has been aligned with Neuroinformatics Platform (SP5) requirements. The Neuroinformatics Platform's access to the data delivered has almost been completed. T1.2.2 has started to work on data integration inside the HBP data space and has provided SP5 with the first dataset of cell distribution (purkinje cells).

In this Task, the data set scheduled to be delivered at the end of the Ramp-Up Phase (M30) comprises three cell distribution maps of the whole mouse brain. By Month 18, T1.2.2 had generated one cell distribution map, in line with plans (see table 1 at the end of this chapter).

The full strategic data set likely to be generated in the long term in this task will feed the Platforms

2.6.2 Data Provenance

The data generated have been captured on site at Pavone's laboratory.

2.7 First Numbers and Distributions of Neurons and Glia

Responsible: T1.2.3 (UPM)

2.7.1 Data Description

- Species: *mus musculus*, GAD67GPF strain
- Brain region: somatosensory cortex
- Origin: *post mortem*
- Data format: tiff image stacks and excel files

In T1.2.3, GAD67GFP mice were used to study the density and distribution of different types of cells in the somatosensory cortex in fixed brains. Data have been delivered as tiff image stacks and excel files. The format of these data has been aligned with Neuroinformatics Platform (SP5) requirements. The Neuroinformatics Platform's access to the delivered data set is currently being established.

In this Task, the generation of maps of inhibitory and excitatory neurons and glia is scheduled to be completed in Month 30. The establishment of automated methods to generate maps of inhibitory and excitatory neurons and glia has been achieved as scheduled. The method was tested and validated and applied in different cortical layers. By Month 18, 35% of the map generation work had been accomplished.

In this Task, the strategic data generated will feed the Platforms.

2.7.2 Data Provenance

The data generated have been captured on site at DeFelipe's laboratory.

2.8 First Morphologies of Neurons and Glia in Hippocampus and Neocortex

Responsible: T1.2.4 (WMC for Neocortex & IEM HAS for Hippocampus)

2.8.1 Data Description

2.8.1.1 Neocortex:

- Species: *mus musculus*, *rattus norvegicus*
- Brain region: neocortex
- Cell type: neurons
- Origin: *post mortem*
- Data format: image stacks [Data files (3d graph); tiff; xls]

The neurons that T1.2.4 has reconstructed have been shared with collaborators in the Brain Simulation Platform (SP6) and the Neuroinformatics Platform (SP5), so that these

data can be included in the relevant databases and the reconstructed neurons can be put into their corresponding sites in a 3D mouse brain atlas.

The Neuroinformatics platform (SP5) currently has access to some but not all of the data delivered so far. In particular, the reconstructed Golgi cells (n = 41) have been delivered to the SP6 team in Geneva via Ms Ying Shi, who works under Sean Hill's supervision.

The data set that T1.2.4 expects to deliver should include 50-70 mouse cell morphologies reconstructed in the neocortex in Month 30. In the neocortex, there are different compositions of neuronal morphologies in different cortical regions. In collaboration with SP6, more than 55 types of excitatory and inhibitory neurons have been classified, which includes almost all neuronal types in the somatosensory cortex of rat. Since Month 11, reconstructions of single pyramidal neurons in a whole brain wide using GFP-mouse-brain image stacks have been carried out.

Data generation up to Month 18 data has been as planned. The reconstruction of 50 neurons from mouse brain has been carried out, including 41 Golgi cells with incomplete structures and nine pyramidal cells with their projections in a whole brain wide (see Table 1 at the end of this chapter).

The full strategic data set likely to be generated in the long term in this task will feed the Platforms.

2.8.1.2 Hippocampus:

- Species: *mus musculus* (8 week-old Black6 mice)
- Brain region: hippocampus
- Cell type: neurons
- Origin: *in vitro*
- Data format: time series txt files; image stack; images

The Neuroinformatics and Brain Simulation Platforms, SP5 and SP6 respectively, are currently being given access to the delivered morphological and electrophysiological data. We have started to share these data with collaborators in SPs 5 and 6, so they can be included in the databases and used in the construction of the unifying hippocampal model. Following extensive consultation, all details of the data transfer have been agreed on, including the final form of the metadata file and the structure of the directories containing all data relevant for the individual reconstructions, as well as the protocol list. We are now ready to upload all of our reconstructed cells.

In T1.2.4, the data set scheduled to be delivered should comprise 110 mouse cell morphologies reconstructed in the hippocampus at the end of the first project phase (Month 30). By Month 18, 23 cells had been fully reconstructed and 19 visualised and scanned (see Table 1 at the end of this chapter). In Months 13—18, T1.2.4 (Hippocampus) has mainly focused on refining and validating the precise measurement of dendritic diameters, and on establishing data formats and the flow of information for depositing data in the databases of the Neuroinformatics Platform.

The full strategic data set likely to be generated in the long run in this task will feed the Platforms.

2.8.2 Data Provenance

The data generated have been captured on site at Markram's and Wang's laboratories for data generated from the neocortex, and at Freund's laboratory for data from hippocampus.

Task	Indicator	Month 6		Month 12		Month 18		Month 24		Month 30	
		Planned	Achieved	Planned	Achieved	Planned	Achieved	Planned	Achieved	Planned	Achieved
T1.1.2	No. of neocortical neuron types with first draft SCTs	0	0	0	0	6	22	6	TBC	6	TBC
T1.1.2	No. of hippocampal neuron types with first draft SCTs	0	0	0	0	0	0	4	TBC	4	TBC
T1.1.2	No. of cerebellum neuron types with first draft SCTs	0	0	0	0	0	0	0	TBC	2	TBC
T1.1.3	No. of brain region synapse proteomes	0	0	0	0	15	42 brain region synapse proteome profiles generated by LC-MS/MS. (7 integral regions of the mouse brain from n=6 mice)	20	TBC	25	TBC
T1.2.1	No. of maps of mouse vasculature	0	0	4	4	6	6	6	TBC	0	TBC
T1.2.2	No. of cell distribution maps of the whole mouse brain	0	0	0	0	1	1	2	TBC	3	TBC
T1.2.3	No. of synaptic maps of individual identified neurons	0	0	0	0	12	12	24	TBC	60	TBC
T1.2.4	No. of mouse cell morphologies reconstructed (hippocampus)	0	0	20	13 cells fully reconstructed & 13 visualised and scanned	50	23 cells fully reconstructed & 19 visualised and scanned	80	TBC	110	TBC
T1.2.4	No. of mouse cell morphologies reconstructed (neocortex)	8	8 (mouse)	20-40	20-40 (mouse) & 30 single PCs and interneurons (rat) & 19 neurons (rat)	30-50	50 neurons (41 Golgi cells with incomplete structures and 9 PCs)	40-60	TBC	50-70	TBC
T1.2.5	No. of axonal projections traced	1	1	4	6 Thalamocortical SS & 1 Thalamocortical visual	12	11 axons	20	TBC	30	TBC
T1.2.5	No. of terminal axon branches and monosynaptic targets identified	4	4	10	10	20	20 targets	30	TBC	40	TBC
T1.2.6	No. of synaptic maps of mouse brain	0	0	0	0	1	1	2	TBC	2	TBC

Table 1: Planned and achieved numerical indicators for the SP1 tasks

3. How Can Platform Developers Provide Feedback on The Data?

3.1 Names and contact details of data providers and users

WP1.1:

- T1.1.2: Chris Ponting, UOXF: chris.ponting@dpag.ox.ac.uk
- T1.1.3: Seth Grant, UEDIN: seth.grant@ed.ac.uk

WP1.2:

- T1.2.1: Bruno Weber, UZH: bweber@pharma.uzh.ch
- T1.2.2:
  - Francesco Pavone, LENS: francesco.pavone@unifi.it
  - Ludovico Silvestri, LENS: silvestri@lens.unifi.it
- T1.2.3:
  - Javier DeFelipe, UPM: defelipe@cajal.csic.es
  - Lidia Alonso-Nanclares, UPM-CSIC: aidil@cajal.csic.es
  - Alberto Muñoz, UPM: amunozc@bio.ucm.es
- T1.2.4:
  - Neocortex: Yun Wang, WMC: yun.wang@tufts.edu
  - Hippocampus:

    Tamas Freund, IEM HAS: freund@koki.hu

    Szabolcs Kali, IEM HAS: kali@koki.hu
- T1.2.5: Francisco Clascá, UAM: francisco.clasca@uam.es
- T1.2.6:
  - Javier DeFelipe, UPM: defelipe@cajal.csic.es
  - Ruth Benavides-Piccione, UPM-CSIC: rbp@cajal.csic.es
  - Ángel Merchán, UPM: amerchan@fi.upm.es

3.2 Types of helpful feedback required by data providers and platforms users

3.2.1 Molecular Section

By means of teleconferencing and email, Douglas Armstrong (SP1), Catherine Zwahlen (SP5), Martin Telefont (SP5), Emily Clark (SP5), Oksana Sorokina (SP1) and David Sterratt (SP1) have agreed:

- A methods ontology to the describe methods by which proteomic and transcriptomic data are collected

- Metadata to describe the data itself.

To test and refine the methods ontology and the metadata, we have extracted methods and data from a number of papers and datasets from the literature, as well as some unpublished data from Seth Grant and Marcia Roy:

- Boyken et al. (2013) Molecular profiling of synaptic vesicle docking sites reveals novel proteins but few differences between glutamatergic and GABAergic synapses. *Neuron* **78**, 285-297[4]
- Wilhelm et al. (2014) Composition of isolated synaptic boutons reveals the amounts of vesicle trafficking proteins. *Science* **344**, 1023-1028[5]
- Biesemann et al. (2013) Proteomic screening of glutamatergic mouse brain synaptosomes isolated by fluorescence activated sorting. *EMBO Journal* **33**, 157-170[6]

  Distler et al. (2014) In-depth protein profiling of the postsynaptic density from mouse hippocampus using data-independent acquisition proteomics . *Proteomics* **14**, 2607-2613[7]

For each paper we have extracted the data tables, and produced a metadata file that describes:

- The animals sampled (species, strain, sex and age)
- The tissues sampled (location and extraction method)
- Derived samples (e.g. Western blots or SDS-Page)
- The type of data described by each table (e.g. Proteins identified), the method by which this data was obtained (e.g. Mascot search using MS data)
- Any post-experiment work carried out, e.g. Mapping tables of proteins onto UNIPROT identifiers.

This data has been deposited in the Neuroinformatics Platform, which will make clear the provenance of the data, including the methods used to map proteomic and transcriptomic data onto standardised identifiers. We will extend this work by extracting data and methods from 31 other publications, which may necessitate small modifications to the metadata, but we are now in a position to deposit data generated in SP1 in the Neuroinformatics Platform.

3.2.2 Cellular Section

In the integration of the data generated, the most critical issue is to work in close collaboration with the Platform users, data providers and Platform developers. Data providers should know how their data are being used by the Platforms and whether they are useful for developers and users. A regular communications flow has been set up to ensure that the technical requirements of the Platforms are matched to the strategic data that are being generated in the SP, to guarantee the effectiveness of the developed tools.

Annex A: Glossary

Term	Explanation
BDA	biotinylated dextran amine (molecular probe)
C57BL6	C57 black 6 is a common inbred strain of laboratory mouse.
GAD67GFP	Mouse strain originally developed by Yuchio Yanagawa. These mice express fluorescent labelling (GFP), specifically in the majority of GABAergic neurons. These neurons can be detected easily by fluorescent microscopy without immunostaining.
LC-MS/MS	Liquid chromatography–mass spectrometry/mass spectometry
LRPNs	Long-Range Projection Neurons
Mesh	A mesh is the extraction of an isosurface that approximates an original element. This process provides an accurate representation of the real shape that can be reconstructed from a stack of images, using specific tools for thresholding and isosurface extraction
Pal-GFP	Green fluorescent protein with a palmitoylation signal
PSD95	Postsynaptic density protein
PSD95GFP	Transgenic mice expressing the enhanced green fluorescent protein (EGFP) aPSD95 protein
Thalamic nuclei DLG	Dorsal lateral geniculate nuclei
Thalamic nuclei LP	Lateral posterior nucleus
Thalamic nuclei Po	Posterior thalamic nuclear group
Thalamic nuclei VPM	Ventral posteriomedial nucleus
VTA projection neurons	Projections of the ventral tegmental area

Annex B: References

[1] Zeisel A, Munoz-Manchado AB, Codeluppi S, Lonnerberg P, La Manno G, Jureus A, et al. Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science. 2015;347(6226):1138-42

[2] Kuramoto E, Furuta T, Nakamura KC, Unzai T, Hioki H, Kaneko T. Two types of thalamocortical projections from the motor thalamic nuclei of the rat: a single neuron-tracing study using viral vectors. Cereb Cortex. 2009 Sep; 19(9):2065-77. doi: 10.1093/cercor/bhn231.

[3] Gong H, Zeng S, Yan C, Lv X, Yang Z, Xu T, Feng Z, Ding W, Qi X, Li A, Wu J, Luo Q. Continuously tracing brain-wide long-distance axonal projections in mice at a one-micron voxel resolution. Neuroimage. 2013 Jul 1; 74:87-98. doi: 10.1016/j.neuroimage.2013.02.005.

[4] Boyken J, Gronborg M, Riedel D, Urlaub H, Jahn R, Chua JJ. Molecular profiling of synaptic vesicle docking sites reveals novel proteins but few differences between glutamatergic and GABAergic synapses. Neuron. 2013;78(2):285-97.

[5] Wilhelm BG, Mandad S, Truckenbrodt S, Krohnert K, Schafer C, Rammner B, et al. Composition of isolated synaptic boutons reveals the amounts of vesicle trafficking proteins. Science. 2014;344(6187):1023-8

[6] Biesemann C, Gronborg M, Luquet E, Wichert SP, Bernard V, Bungers SR, et al. Proteomic screening of glutamatergic mouse brain synaptosomes isolated by fluorescence activated sorting. The EMBO journal. 2014;33(2):157-70.

[7] Distler U, Schmeisser MJ, Pelosi A, Reim D, Kuharev J, Weiczner R, et al. In-depth protein profiling of the postsynaptic density from mouse hippocampus using data-independent acquisition proteomics. Proteomics. 2014;14(21-22):2607-13

Other Reading

Boyken J, Gronborg M, Riedel D, Urlaub H, Jahn R, Chua JJ. Molecular profiling of synaptic vesicle docking sites reveals novel proteins but few differences between glutamatergic and GABAergic synapses. Neuron. 2013;78(2):285-97.

Keller PJ, Ahrens MB. Visualizing whole-brain activity and development at the single-cell level using light-sheet microscopy. Neuron. 2015 Feb 4; 85(3):462-83. doi: 10.1016/j.neuron.2014.12.039

Kuramoto E, Furuta T, Nakamura KC, Unzai T, Hioki H, Kaneko T. Two types of thalamocortical projections from the motor thalamic nuclei of the rat: a single neuron-tracing study using viral vectors. Cereb Cortex. 2009 Sep; 19(9):2065-77. doi: 10.1093/cercor/bhn231.