

Description of Additional Supplementary Files

Supplementary Data Legends

Supplementary Data 1. Synaptic hippocampal proteomes characterised in this study.

Sheet #1: Proteins identified by our proteomics workflow from a biochemical preparation of postsynaptic density fractions from total hippocampus.

Sheet #2: Reference synaptic proteome used in this study. Produced by combining proteins in sheet #1 with those previously identified in the PSDII fraction published by Distler et al¹.

Sheet #3: List of all proteins identified by the Scaffold software from MS/MS data in synaptic fractions from the three hippocampal layers. Proteins common with the reference proteome are indicated. Proteins identified by the Progenesis software with at least two unique peptides are also indicated.

Sheet #4: Proteins identified by scaffold in only one of the three synaptic types studied.

Supplementary Data 2. Analysis of protein abundance and differential protein expression between synaptic types.

Sheet#1: Protein abundance data generated by MSqROB^{2,3} from peptide abundance data.

Sheet#2: Left of black bar, Proteins with statistically highest expression in one synaptic type. Statistics, one-way ANOVA. FDR correction for multiple testing was performed. Log2 of Fold Chance (FC) and corrected p-values (q-value) are provided. Right of black bar, Protein abundance differences between pairs of synaptic types. Statistics, Student's T-test. FDR correction for multiple testing was performed. Log2 of Fold Chance (FC) and corrected p-values (q-value) are provided.

Supplementary Data 3. Comparative analysis of protein and RNA expression data from proteins differentially expressed in synapses from the trisynaptic loop.

Sheet#1: Allen Brain Atlas (ABA) *in situ* hybridization (ISH) data was manually inspected for each of the 283 proteins showing differential expression in one of the synapses from the trisynaptic loop (columns E to G). We determined in how many of the 4 brain regions forming the synapses from the trisynaptic loop (Entorhinal cortex Layer II, dDG, CA3 and CA1) was ISH data highest (columns H to K). We compared this information with our

proteomics data and established if expression levels were concordant or not with protein levels at synapses (column M).

Sheet#2: Synaptic genes significantly up or down regulated in excitatory neurons from one of the 4 brain subregions constituting the trisynaptic circuit of the hippocampus: CA1 (dorsal), CA3 (dorsal), dentate gyrus and entorhinal cortex. RNA sequencing data taken from the Allen Brain Cell atlas (ABCA). The ABCA distinguishes dorsal from ventral neurons in the CA1 and CA3 subfields. As the proteomics data was generated from the dorsal hippocampus, we selected dorsal neurons from the ABCA for this analysis. Log2 fold changes and p-values are indicated. Statistical analysis of RNA abundance data performed using the Seurat R package, see methods. Statistical Test, two-sided Wilcoxon Rank Sum test. Correction for multiple testing performed using the Benjamini-Hochberg procedure.

Supplementary Data 4. Signalling pathways and GO terms significantly overrepresented in proteins with highest expression in one synaptic type.

Sheet#1: The analysis with pathfindR retrieved the following Signalling Pathways as significantly overrepresented amongst protein with highest expression in each synaptic type. Signalling pathways were retrieved from the following databases: Reactome, KEGG and Wikipathways (WP). Fold enrichments are provided, these are calculated as the number of proteins observed in a pathway or term relative to the number expected by chance. The pathfindR metrics occurrence, support, lowest and highest p-values, cluster and status are also provided. Protein with no expression difference between synapses belonging to each pathway are also shown (column K). Proteins from each pathway with highest expression in one synaptic type are indicated (column L). Statistical test, one-sided hypergeometric tests, correcting for multiple testing with the Bonferroni method (see methods regarding pathfindR analysis workflow).

Sheet#2: The analysis with pathfindR retrieved the following GO terms as significantly overrepresented amongst protein with highest expression in each synaptic type. GO terms from the following domains were investigated: Molecular Function (GOMF), Biological Process (GOBP) and cellular component (GOCC). Fold enrichments are provided, these are calculated as the number of proteins observed in a pathway or term relative to the number expected by chance. The pathfindR metrics occurrence, support, lowest and highest p-values, cluster and status are provided. Proteins with no expression difference between synapses belonging to each term are shown (Col. K). Proteins from each term with highest expression in one synaptic type are indicated (Col. L). Statistical test, one-sided hypergeometric tests, correcting for multiple testing with the Bonferroni method (see methods regarding pathfindR analysis workflow).

Sheet#3: Summary of pathways and terms identified as 'Representative' for networks (clusters) of proteins with highest expression in different synaptic types, as determined by PathfindR. A representative pathways or term is the one with the lowest p-value amongst those identified for a protein network.

Supplementary Data 5. Genes coding for synaptic proteins that have differential RNA expression levels between neuronal classes.

Sheet#1: Table with the number of genes found with a statistically significant up- or down-expression in each neuronal class.

Sheet#2: List of genes significantly up- or down-regulated in each class. The ratio of genes up vs. down-regulated is also provided. Statistical analysis of RNA abundance data performed using the Seurat R package, see methods. Statistical Test, two-sided Wilcoxon Rank Sum test. Correction for multiple testing performed using the Benjamini-Hochberg procedure.

Sheet#3: List of genes significantly up-regulated in one or two classes. Statistical analysis of RNA abundance data performed using the Seurat R package, see methods. Statistical Test, two-sided Wilcoxon Rank Sum test. Correction for multiple testing performed using the Benjamini-Hochberg procedure.

Supplementary Data 6. Genes coding for synaptic proteins that have differential RNA expression levels between neuronal types of the same class.

Sheet#1: Table with the number of genes found with a statistically significant up- or down-expression between neuronal types of each class. Neuronal type names as previously published.

Sheets#2, 4, 6, 8,10, 12, 14 and 16: Lists of genes significantly up- or down-regulated between neuron types of each of the eight classes investigated. Statistical analysis of RNA abundance data performed using the Seurat R package, see methods. Statistical Test, two-sided Wilcoxon Rank Sum test. Correction for multiple testing performed using the Benjamini-Hochberg procedure.

Sheets#3, 5, 7, 9,11, 13, 15 and 17: List of genes significantly upregulated in one or two neuronal types within each class. Statistical analysis of RNA abundance data performed using the Seurat R package, see methods. Statistical Test, two-sided Wilcoxon Rank Sum test. Correction for multiple testing performed using the Benjamini-Hochberg procedure.

Sheet#18: Summary table of genes coding for synaptic proteins significantly upregulated in one or two neuronal types.

Supplementary Data 7. Signalling pathways and GO terms enriched among genes upregulated in different neuronal types.

Sheet#1: List of representative pathways from the databases Reactome, KEGG and Wikipathways identified by pathfindR for synaptic genes upregulated in different neuronal types. Statistical test, one-sided hypergeometric tests, correcting for multiple testing with the Bonferroni method (see methods regarding pathfindR analysis workflow).

Sheet#2: List of representative GO terms identified by pathfindR for synaptic genes upregulated in different neuronal types. Statistical test, one-sided hypergeometric tests, correcting for multiple testing with the Bonferroni method (see methods regarding pathfindR analysis workflow).

Supplementary Data 8. Top 1000 genes contributing to the transcriptomics-based classification of excitatory neurons and analysis of the signalling pathways and GO terms associated to them.

Sheet#1: List of the 1000 proteins mostly contributing to the classification of excitatory neurons into classes, as determined by the Random Forest method.

Sheet#2: List of the 1000 proteins mostly contributing to the classification of excitatory neurons into types, as determined by the Random Forest method.

Sheet#3: Representative terms (Signalling Pathways and GO terms) identified by pathfindR from the synaptic genes among the top 1000 most contributing to the classification of excitatory neurons. Statistical test, one-sided hypergeometric tests, correcting for multiple testing with the Bonferroni method (see methods regarding pathfindR analysis workflow).

Sheet#4: Representative terms (Signalling Pathways and GO terms) identified by pathfindR from the non-synaptic genes among the top 1000 most contributing to the classification of excitatory neurons. Statistical test, one-sided hypergeometric tests, correcting for multiple testing with the Bonferroni method (see methods regarding pathfindR analysis workflow).

Sheet#5: Summary of the synaptic and non-synaptic terms used in Figure 6g and h.

Supplementary Video 1. Manual dissection of hippocampal subfields.

Supplementary Code Files:

Supplementary_Code_1_Iteration_Classes.R: R script to iterate the statistical analysis performed with Seurat to identify genes differentially expressed between neuronal classes.

Supplementary_Code_2_Iteration_Types.R: R script to iterate the statistical analysis performed with Seurat to identify genes differentially expressed between neuronal types.

Supplementary_Code_3_Analysis_Classes.R: R script to generate data tables and graphs for genes differentially expressed between neuronal classes. This script also includes a quality control test to validate differentially expressed genes.

Supplementary_Code_4_Analysis_Types.R: R script to generate data tables and graphs for genes differentially expressed between neuronal Types. This script also includes a quality control test to validate differentially expressed genes.

Supplementary_Code_5_Split_Types.R: R script to obtained data from a subset of neuronal types from the entire transcriptomic database provided by the ABCA.

Supplementary_Code_6_pathfindR_Proteomics.Rmd: R script to perform the pathfindR analysis and to generate the heatmaps from the proteomics data.

Supplementary_Code_7_pathfindR_Classes.R: R script to perform the pathfindR analysis and to generate the heatmaps from transcriptomics data of neuronal classes (ABCA).

Supplementary_Code_8_pathfindR_Types.R: R script to perform the pathfindR analysis and to generate the heatmaps from transcriptomics data of neuronal types (ABCA).

Supplementary_Code_9_Random_Forest.ipynb: Python code to perform the Random Forest analysis on transcriptomic data from the ABCA.