Synaptic proteome diversity is primarily driven by gene regulation of glutamate receptors and their regulatory proteins

Authors:

Rita Reig-Viader[1,2*], Diego del Castillo-Berges[1,2*], Albert Burgas-Pau[1,2,3,4*], Daniel Arco-Alonso[1*], David Ramos-Vicente[1,5], Carlos Sindreu[6], Àlex Bayés[1,2#]

1. Molecular Physiology of the Synapse Laboratory, Institut de Recerca Sant Pau (IR Sant Pau), Barcelona, Spain.
2. Universitat Autònoma de Barcelona (UAB), Bellaterra (Cerdanyola del Vallès), Spain.
3. Unitat mixta d'Investigació IRTA-UAB en Sanitat Animal. Centre de Recerca en Sanitat Animal (CReSA), Bellaterra (Cerdanyola del Vallès), Spain.
4. IRTA. Programa de Sanitat Animal. Centre de Recerca en Sanitat Animal (CreSA), Bellaterra, 08193, Catalonia. Spain.
5. Neurodegenerative Diseases Research Group, Vall d'Hebron Research Institute, Centre for Networked Biomedical Research on Neurodegenerative Diseases (CIBERNED), Barcelona, Catalonia, Spain.
6. Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain.

* These authors contributed equally to the work

Corresponding author:

Àlex Bayés, Molecular Physiology of the Synapse Laboratory, Institut de Recerca Sant Pau (IR Sant Pau), Barcelona, Spain., C/Sant Quintí, 77-79, 08041 Barcelona, Spain.

Email: abayesp@santpau.cat

Abstract

Electrophysiological features of excitatory synapses vary widely throughout the brain, granting neuronal circuits the ability to decode and store diverse patterns of information. Synapses formed by the same neurons have similar electrophysiological characteristics, belonging to the same type. However, these are generally confined to microscopic brain regions, precluding their proteomic analysis. This has greatly limited our ability to investigate the molecular basis of synaptic physiology. Here we introduce a procedure to characterise the proteome of individual synaptic types. We reveal a remarkable proteomic diversity among the synaptic types of the trisynaptic circuit. Differentially expressed proteins participate in well-known synaptic processes, controlling the signalling pathways preferentially used among diverse synapses. Noteworthy, all synaptic types differentially express proteins directly involved in the function of glutamate receptors. Moreover, neuron-specific gene expression programs would participate in their regulation. Indeed, genes coding for these proteins exhibit such distinct expression profiles between neuronal types that they greatly contribute to their classification. Our data is an important resource for exploring the molecular mechanisms behind electrophysiological properties of different hippocampal synaptic types. Our combined analysis of proteomics and transcriptomics data uncovers a previously unrecognised neuron-specific transcriptomic control of synaptic proteome diversity, directed towards the regulation of glutamate receptors and their regulatory proteins.

Keywords: Synaptic type, proteomics, proteome diversity, transcriptomics, laser-capture microdissection, hippocampus, trisynaptic circuit, glutamate receptors, gene regulation.

Introduction

Proteomics research performed on synaptic biochemical preparations has established a very comprehensive catalogue of proteins that play a role in synaptic biology[1–7]. This central advance in brain research has nevertheless been limited by the requirements of biochemical fractionation procedures and the sensitivity limitations of proteomics methods. These have imposed to work with relatively large brain areas, such as the hippocampus or neocortex[6,8–11]. Yet, these brain samples are not homogenous, containing many different synaptic types that are analysed together[12]. Accordingly, proteomics research has uncovered the composition of the average, or the prototypical, synapse in a given sample. However, to understand the molecular mechanisms orchestrating the functional states that a synapse can take, it is imperative to investigate individual synaptic types. This is arguably the most important technical hurdle to precisely elucidate the molecular mechanisms behind synaptic function, with implications on information processing and cognition.

Synaptic types can be defined in different ways, for instance they can be chemical or electrical; they can also be defined based on their neurotransmitter content, the pair of neurons forming them or as recently shown, according to the expression patterns of key scaffolding molecules[13,14]. In the present work a synaptic type refers to that formed by a specific pair of pre- and post-synaptic neurons. This is because there is an extensive electrophysiological literature showing that synapses defined by connectivity have different functional properties[12,15–17]. A paradigmatic example is to be found in the hippocampus, where functional differences between CA3-CA1 and DG-CA3 glutamatergic synapses are prominent[17].

Several methodological approaches have appeared in recent years to get closer to the final goal of isolating individual synaptic types or even individual synapses. All of them have been performed in mice and rely on genomic manipulations. Some of these approaches used fluorescently tagged proteins to sort synaptosomal preparations[18–22]. These methods have allowed to investigate glutamatergic neurons in large brain regions, or to investigate the cell-surface proteome of mossy fibre synapses in CA3[20]. Other approaches took advantage of proximity labelling methods to define the proteome of inhibitory synapses or the synaptic cleft[23–25]. More recently, confocal imaging studies in mice expressing three of the four proteins in the Psd95 family tagged with different fluorophores, provided a glimpse at the daunting molecular diversity that excitatory synapses could have, without losing anatomical information[14,26]. These cutting-edge studies are starting to uncover the molecular diversity among synapses, that could only be suspected until now. Nevertheless, these approaches are not fit to explore the large proteomic landscapes of local synaptic types, and have low translational power, as they cannot be used in human samples. So far, research on synaptic

proteome diversity has not been able to provide a general framework or a set of general principles to explain this variability.

To start addressing the molecular diversity between types of glutamatergic synapses, we leveraged on the topographical organization of the hippocampus. Which contains one of the best studied neuronal circuits in the brain, the trisynaptic circuit. This is formed by three types of excitatory synapses that localize to anatomically different layers. Importantly, each of these layers contains mostly one synaptic type[17,27–29]. In this circuit, the first synapse is made between the axons of layer II neurons from the entorhinal cortex and the dendrites of granular cells in the dentate gyrus molecular layer (EC-DG). In turn, granule cell axons give rise to the mossy fibres that contact the proximal dendrites of CA3 pyramidal cells in the striatum lucidum (DG-CA3). Finally, the third synapse is formed by axons leaving CA3 neurons and contacting the proximal dendrites of CA1 pyramidal neurons in the striatum radiatum (CA3-CA1). Electrophysiological studies have demonstrated that these synapses have different functional characteristics, displaying unique synaptic transmission and plasticity features[30].

To reveal the proteomic diversity between glutamatergic synaptic types, we developed a high-yield procedure that allows to characterise their proteome. Using this method, we uncovered the proteomic diversity of the synaptic types forming the trisynaptic circuit of the dorsal hippocampus. We also investigated expression differences of genes coding for synaptic proteins in 55 neuronal types from the hippocampus and subiculum. Together our proteomics and transcriptomics analysis indicate that abundance differences in glutamate receptors and the proteins that regulate them are common drivers of proteome variability across synaptic types and that neuron-specific gene expression mechanisms participate in this regulation.

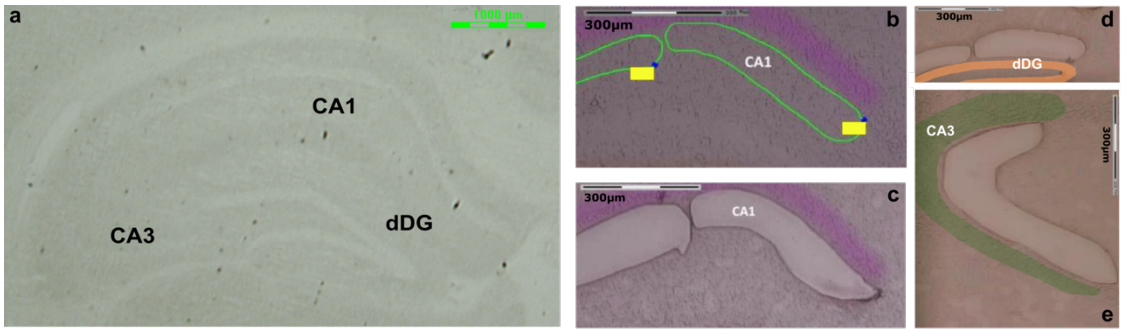
Results

Development of a procedure to obtain synaptic proteins from microscopic samples.

To increase the anatomical resolution of synapse proteomics we have developed a procedure to extract synaptic proteins from microscopic brain regions. This method combines laser-capture microdissection (LCM) with enhanced extraction and recovery of synaptic proteins. We applied this procedure to perform deep proteomic profilings of the synaptic types constituting the trisynaptic circuit from the dorsal hippocampus.

In this procedure forebrains are dissected and rapidly snap-frozen prior to cryosectioning. Brains cannot be chemically fixed, as this negatively interferes with later proteomic analysis. We stablished maximum section thickness for effective LCM cutting to be 10 μm. Microdissection was performed in coronal slices encompassing the first 500 μm of the dorsal hippocampus (Suppl. Fig1a). As the pyramidal and granular layers, which contain cell bodies, can be visually distinguished (Fig. 1a), they can be excluded, collecting only the synaptic-rich neuropile (Fig.1b-c, for an example at CA1). By dissecting fragments of 100 μm in width it is possible to have control over the hippocampal layer collected (Suppl. Fig 1b-c). From the dentate gyrus we obtained the Molecular Layer (ML, Fig. 1d), from CA3 we dissected the Stratum Lucidum (SL, Fig. 1e) and from CA1 the Stratum Radiatum (SR, Fig. 1c). The higher translucidity of the SL helped in localizing and collecting this layer.

Extracting synaptic proteins from the microscopic amounts of tissue collected by LCM is extremely challenging. To cope with this limitation, we developed a procedure designed to minimize sample manipulation, which increases sample loss, while maximizing recovery of synaptic proteins. This procedure takes advantage of the selective solubility of synaptic structures to the detergent Triton X-100, such as the postsynaptic density (PSD), the active zone (AZ) or the extracellular matrix of the synaptic cleft[5]. First, microdissected tissue is accumulated in a solution containing 1% Triton X-100 (Fig. 1f-h). Next, neuropile fragments are subjected to a three-step treatment, a brief bath sonication, a mild thermal shock at 35C in agitation, and a second sonication step. This procedure fully disperses neuropile fragments and maximises the effect of the detergent, while preserving protein integrity and avoiding sample manipulation. A final centrifugation allows to collect Triton-insoluble proteins (Fig. 1i).



f

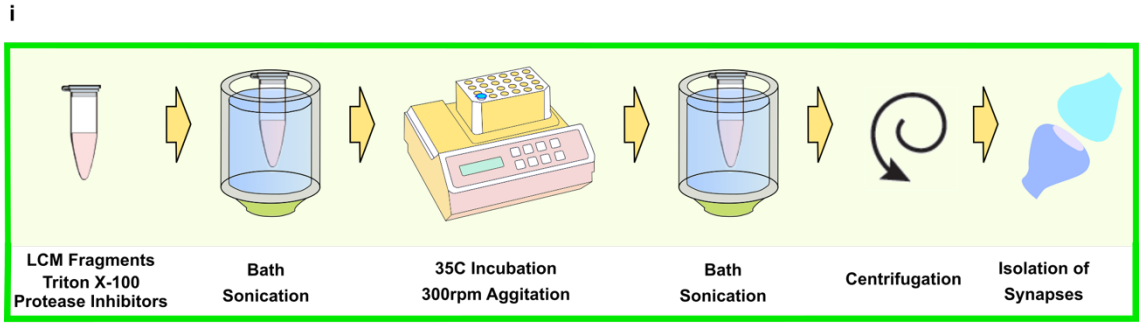
Sample	Total mm²	Sections
1	95,93	250
2	91,73	240
3	100,64	270

g

Sample	Total mm²	Sections
1	92,25	250
2	103,56	320
3	96,02	270

h

Sample	Total mm²	Sections
1	105,65	280
2	94,35	250
3	90,76	240



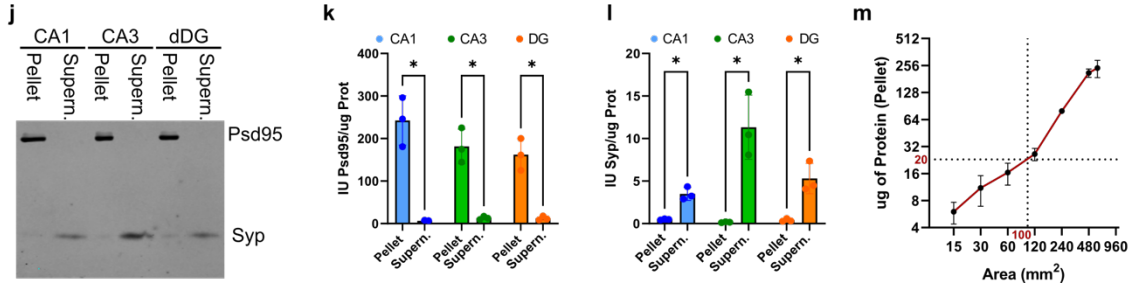


Figure 1. Effective separation of proteins from each of the synapses constituting the trisynaptic circuit using laser-capture microdissection and biochemical processing of hippocampal layers.

a. Brightfield image showing the hippocampus in a coronal section of the dorsal mouse brain used for laser-capture microdissection (LCM). Image taken with the microscope used for LCM. Note that CA1/CA3 pyramidal layer and dDG granular layer are visible in hippocampal subfields. Scale bar 1000µm.

b. Brightfield image of the CA1 subfield before microdissection. Pyramidal layer is highlighted in purple. The green line marks the area that will be microdissected. Microdissected fragments had a width of approximately 100µm, thus only collecting neuropile from the Stratum Radiatum layer. Scale bar 300µm.

c. Brightfield image of the CA1 subfield from the section shown in (b) after LCM. The pyramidal layer, highlighted in purple, is not collected. Scale bar 300µm.

d. Brightfield image of the dorsal dentate gyrus after LCM. Microdissected fragments had a width of approximately 100µm, which allowed the specifically collect neuropile from the Molecular Layer. The granular layer is highlighted in orange. Scale bar 300µm.

e. Brightfield image of the CA3 subfield after LCM. Microdissected fragments had a width of approximately 100µm, which allowed to collect neuropile from the Stratum Lucidum. The pyramidal layer is highlighted in green. Scale bar 300µm.

f. Total area (mm²) microdissected and number of brain sections collected for each of the three biological replicas analysed by proteomics of the dDG.

g. Total area (mm^2) microdissected and number of brain sections collected for each of the three biological replicas analysed by proteomics of the CA3.

h. Total area (mm^2) microdissected and number of brain sections collected for each of the three biological replicas analysed by proteomics of the CA1.

i. Outline of the procedure used to enrich neuropile samples collected with LCM in synaptic proteins.

j. Immunoblot of 1% Triton X-100 insoluble (Pellet) and soluble (Supern.) fractions obtained from the three hippocampal layers investigated. Proteins analysed are Psd95, a postsynaptic marker, and Synaptophysin (Syp) a synaptic vesicle marker.

k. Bar plot of Psd95 presenting relative protein abundance as determined by immunoblot in 1% Triton X-100 soluble (Supern.) and insoluble (Pellet) fractions from the three hippocampal layers investigated. IU: intensity units. Statistics, Two-way ANOVA and Fisher's LSD post-hoc test, * $p < 0.05$.

l. Bar plot of Synaptophysin (Syp) presenting relative protein abundance as determined by immunoblot in 1% Triton X-100 soluble (Supern.) and insoluble (Pellet) fractions from the three hippocampal layers investigated (Blue, CA1; Green, CA3 and Orange dDG). IU: intensity units. Statistics, Two-way ANOVA and Fisher's LSD post-hoc test, * $p < 0.05$.

m. Micrograms of protein recovered in 1% Triton X-100 pellets per area of microdissected neuropile. To obtain 20μg of protein in insoluble fractions $100mm^2$ of neuropile have to be microdissected.

To evaluate the efficacy of this procedure, we assayed samples by immunoblot against proteins known to be mostly soluble (Synaptophysin, Syp) or insoluble (Psd95) to Triton X-100. Over 90% of the Psd95 signal was detected in pellets (Fig. 1k). Conversely, the same proportion of Syp was in supernatants (Fig.1l). Remarkably, no difference in Psd95 abundance was observed in pellets between samples (two-way ANOVA), indicating that the procedure had a similar efficiency in all hippocampal layers.

As these samples contain very little protein, standard approaches for protein quantification cannot be used. Protein concentration was determined by electrophoresis, using as internal calibration standards hippocampal synaptic preparations accurately quantified (Suppl. Fig. 2a,b). Using this approach, we determined that insoluble fractions contain approximately 20% of all protein in the tissue (Suppl. Fig. 2c), indicating that proteins in these fractions were concentrated 4-5 times. We also tested different extraction buffers to investigate if we could improve the efficiency of the procedure. Using a RIPA buffer we found that the amount of protein recovered in pellets was significantly smaller (Suppl. Figure 2d,e), yet this was at the expense of solubilizing a larger proportion of both Psd95 and Syp (Suppl. Fig. 2f-g). Indicating that more synaptic components were lost in the soluble fraction. On the other hand, increasing Triton concentration to 2% did not improved protein yield (Suppl. Figure 2e). Neither RIPA nor 2% Triton showed improved performance over 1% Triton X-100, which remained as the buffer of choice. Finally, we established how much protein was recovered in pellets per area of microdissected neuropile, this was important to keep LCM time to a minimum. We determined that for each $100mm^2$ of neuropile we obtained approximately 20μg of triton insoluble protein (Fig. 1m). This was sufficient for our proteomics analysis, which routinely require 10μg of protein or less.

Deep proteomic coverage of synaptic types from the trisynaptic circuit reveals high similarity at the composition level.

Using the above procedure, we obtained biological triplicates of synaptic preparations from the layers of the trisynaptic circuit and subjected them to an established proteomics workflow[31]. MS/MS data was examined with Scaffold-DIA (Proteome Software), to identify protein specimens and Progenesis QI (Waters) for high-sensitive peptide quantification (Fig. 2a). Peptide abundance was normalized by the average abundance of peptides from 14 synaptic scaffolding proteins (see methods). This allowed to correct for differences in: i) synaptic yield between preparations and ii) synaptic density between layers. Finally, MsqROB[32,33] was used to identify proteins differentially expressed between synaptic types.

The proteomic dataset obtained from microdissected tissue was benchmarked against a reference proteome. This was generated from the combination of two proteomes of hippocampal synaptic fractions prepared using standard procedures (Suppl. Fig. 2i)[6]. We produced the one of these datasets and the other had been previously published[8] (Fig. 2b and Suppl. Table 1). Proteins detected in LCM samples but absent from the reference proteome were discarded as potential contaminants (Fig. 2a and Suppl. Table 1). Thus, initially Scaffold identified 2905 proteins from microdissected samples, of which 628 were discarded after benchmarking. Of the remaining 2277 proteins Progenesis provided quantitative data with at least 2 unique peptides for 2014 proteins, this being the final dataset investigated (Fig. 2a and Suppl. Table 1).

We next confirmed that our method was able to retrieve proteins from distinct synaptic locations. Using the SynGO database[2] to assign subsynaptic locations onto our dataset, we found that it was enriched in many of them, both pre- and postsynaptically (Fig. 2c). As a matter of fact, pre- and postsynaptic proteins were similarly enriched. The presence of presynaptic proteins in our preparations was confirmed by immunoblot (Suppl. Fig. 2f,j). Thus, our approach provides a wide view into the synaptic proteome.

A small number of proteins were identified only in one synaptic type (CA3-CA = 29, DG-CA3 = 68 and EC-DG = 52, Suppl. Table 1). Potentially these proteins could be very interesting, as they might be markers of synaptic types. Nevertheless, most of them (86%) could only be identified in one of the three replicates, and their abundance was very low (mean 3.45 peptides/protein, compared with 43 peptides/protein for the whole set). Thus, we decided to exclude these molecules from subsequent analysis. Our data indicates that few proteins, if any, will be unique to one synaptic type in the trisynaptic loop. Which means that, at the qualitative level, the molecular machines operating at, otherwise functionally different synaptic types, are virtually identical.

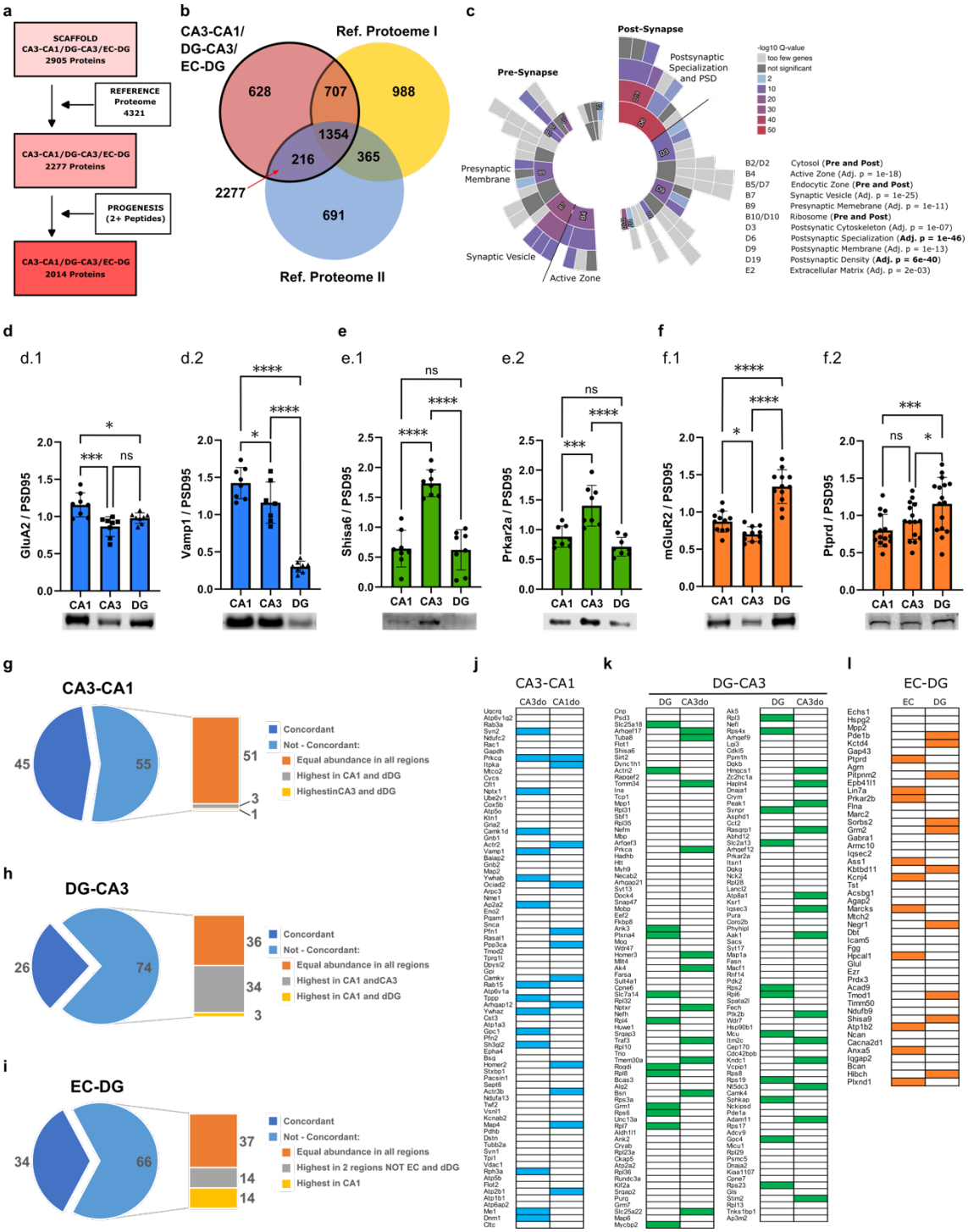


Figure 2. Proteomics profiling of synaptic fractions from the trisynaptic loop identifies proteins differentially expressed in each synaptic type.

a. Steps involved and informatic tools used in the analysis of mass-spectrometry data to establish the final list of proteins in synapses from the trisynaptic loop.

b. Venn diagram showing protein overlap between proteins identified in synaptic fractions from the trisynaptic loop and two hippocampal reference proteomes that use established density gradient ultracentrifugation methods to isolate synaptic fractions. In total 2277 proteins from the LCM dataset overlap with reference proteomes. SR: stratum radiatum from CA1, SL, stratum lucidum from CA3 and ML, molecular layer from dDG. Ref. proteome I, generated in this study; and Ref. Proteome II corresponds with PSDII proteome as defined by Distler et al[8].

c. Sunburst plot showing SynGO Cellular Component terms enriched among proteins identified in synapses from the trisynaptic loop. Note that not only PSD-related locations are found significantly

enriched. Cellular locations corresponding with many other synaptic structures, such as the active zone, synaptic vesicles, endocytic zone, cytosol or even the extracellular matrix, are also represented.

d. Bar plot presents relative abundance of Glua2 (d.1) and Vamp1 (d.2) determined by immunoblot in synaptic fractions isolated from CA1, CA3 and DG hippocampal subfields. A representative immunoblot image is shown. Statistical test used, one-way ANOVA, post-hoc Fisher's LSD test, * $p < 0.05$, *** $p < 0.001$, **** $p < 0.0001$.

e. Bar plot presents relative abundance of Shisa6 (e.1) and Prkar2a (e.2) determined by immunoblot in synaptic fractions isolated from CA1, CA3 and DG hippocampal subfields. A representative immunoblot image is shown. Statistical test used, one-way ANOVA, post-hoc Fisher's LSD test, *** $p < 0.001$, **** $p < 0.0001$.

f. Bar plot presents relative abundance of mGluR2 (f.1) and Ptprd (f.2) determined by immunoblot in synaptic fractions isolated from CA1, CA3 and DG hippocampal subfields. A representative immunoblot image is shown. Statistical test used, one-way ANOVA, post-hoc Fisher's LSD test, * $p < 0.05$, *** $p < 0.001$, **** $p < 0.0001$.

g. Percentage of proteins with highest expression in CA3-CA1 synapses with concordant or discordant RNA expression levels. RNA data obtained from *in situ* hybridization studies deposited in the Mouse Brain Atlas (Allen Brain Map).

h. Percentage of proteins with highest expression in DG-CA3 synapses with concordant or discordant RNA expression levels. RNA data obtained from *in situ* hybridization studies deposited in the Mouse Brain Atlas (Allen Brain Map).

i. Percentage of proteins with highest expression in EC-DG synapses with concordant or discordant RNA expression levels. RNA data obtained from *in situ* hybridization studies deposited in the Mouse Brain Atlas (Allen Brain Map).

j. Proteins with highest expression in CA3-CA1 synapses that also present increased RNA levels in excitatory neurons from the dorsal CA3 (CA3do, left column) or the dorsal CA1 (CA1do, right column) are indicated with a blue box. An empty box denotes no difference at the RNA level. RNA data obtained from single cell RNA sequencing data generated by the Allen Brain Cell atlas[34].

k. Proteins with highest expression in DG-CA3 synapses that also present increased RNA levels in excitatory neurons from the dentate gyrus (DG, left column) or the dorsal CA3 (CA3do, right column) are indicated with a green box. An empty box denotes no difference at the RNA level. RNA data obtained from single cell RNA sequencing data generated by the Allen Brain Cell atlas[34].

l. Proteins with highest expression in EC-DG synapses that also present increased RNA levels in excitatory neurons from the Entorhinal cortex (EC, left column) or the dentate gyrus (CA1do, right column) are indicated with an orange box. An empty box denotes no difference at the RNA level. RNA data obtained from single cell RNA sequencing data generated by the Allen Brain Cell atlas[34].

Gene expression contributes to synaptic proteome variability

The above data implied that quantitative, rather than qualitative, variation drives functional diversity across synapse types. To identify differentially expressed synaptic proteins we used a ridge regression method designed to analyse peptide abundance data acquired by label-free mass spectrometry[32,33]. This approach retrieved a total of 283 proteins, 14% of all, significantly overexpressed in one synaptic type (Suppl. Fig. 3a and Suppl. Table 2). Of these, 78 were from CA3-CA1 synapses, 157 from DG-CA3 synapses and 48 from EC-DG synapses. To validate our proteomics results we manually dissected acute hippocampal slices (Supplementary Video 1), isolated synaptic proteins and performed immunoblot analysis on two highly expressed proteins per layer (Fig. 2d-f). Importantly, the results validated the differential enrichment of all proteins examined.

To investigate if differences in gene expression underlie proteomic changes, we analysed *in situ* hybridization (ISH) data from the Allen Mouse Brain Atlas[35] (Suppl. Table 3 and

methods). Proteomic and ISH data were considered concordant if an upregulated protein showed highest RNA expression in the pre- and/or postsynaptic neurons forming it (e.g. for a CA3-CA1 protein, we would consider ISH data in pyramidal layers from CA3 and CA1). On average, the concordance between RNA and protein expression was 35%, indicating that only a fraction of the proteomic variability between synaptic types is due to gene expression (Fig. 2g-i). To confirm this, we retrieved data from single-cell RNA sequencing (scRNAseq) of excitatory neurons in dorsal CA1, dorsal CA3, DG and entorhinal cortex, from the Allen Brain Cell Atlas (ABCA)[34] and identified upregulated genes (Fig. 2j-l, Suppl Table 3). Again, the concordance of protein and RNA data was around 35%. Hence, transcriptomic mechanisms have a role in defining synaptic proteome variability.

Differentially expressed proteins introduce high diversity in the molecular mechanism operating at individual synaptic types

To investigate the biological functions related to proteins with highest expression in one synaptic type, we performed enrichment analysis of signalling pathways[36–38] and GO terms[39,40] using the pathfindR tool[41]. pathfindR constructs protein-protein interaction networks and maps enriched terms onto them. Using hierarchical clustering and pairwise kappa statistics, pathfindR identifies one 'Representative' term in each network.

We first clarified if a small number of proteins were responsible for a large proportion of enriched terms, a common bias with pathway enrichment analysis[42,43]. Yet this was not the case, as the ratio of enriched terms per protein was low (Suppl. Fig. 3b) and the proportion of proteins contributing to terms was high (Suppl. Fig. 3c) in all synaptic types. Importantly, most enriched pathways (75%) and GO terms (96%) were found only in one synaptic type (Suppl. Fig. 3d,e), thus, effectively informing about their unique functional properties. Only 5 terms were enriched in all synaptic types. These were strongly related to synaptic function and included transmission across chemical synapses, postsynaptic signalling, actin cytoskeleton and cell adhesion (Suppl. Fig. 3f and Suppl. Table 4).

While CA3-CA1 and DG-CA3 synapses shared several functional categories, none was found between CA3-CA1 and EC-DG synapses, and only 3 between DG-CA3 and EC-DG synapses (Suppl. Figure 3e,f, Fig. 3 and Suppl. Table 4). Suggesting that CA3-CA1 and DG-CA3 synapses have a higher degree of similarity regarding their underpinning molecular mechanisms. Of note, the GOCC term 'Schaffer collateral CA1 synapse', appeared enriched in proteins from CA3-CA1 and DG-CA3 synapses. Among the pathways common to these synapses we identified well-known synaptic processes, such as signalling via calcium, through Ras and Rho GTPases or trans-synaptic signalling via BDNF, Ephrins and Semaphorins (Fig. 3).

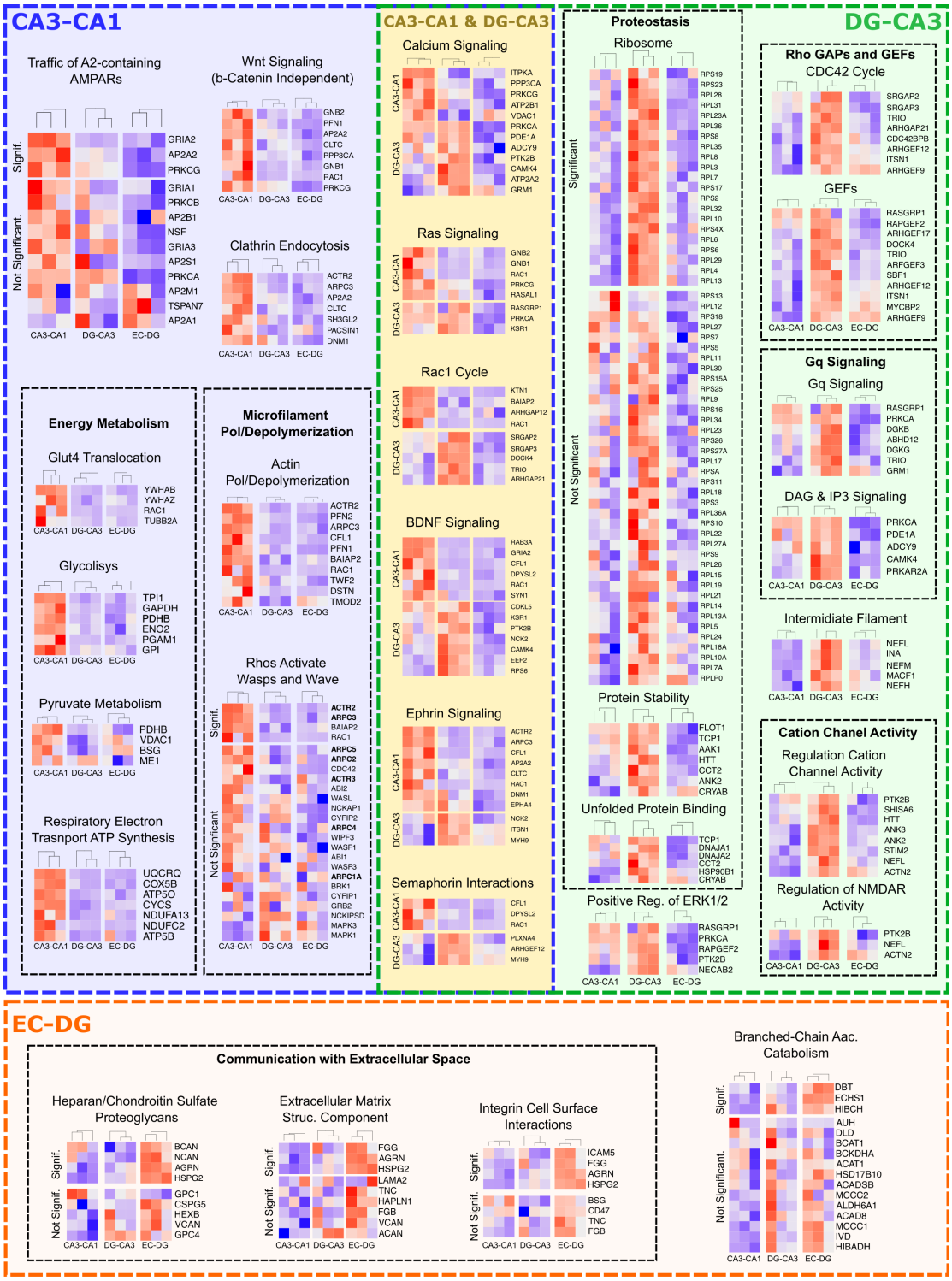


Figure 3. Signalling pathways differentially regulating function in each synaptic type from the trisynaptic circuit.

Signalling pathways overrepresented amongst DE proteins in each synapse of the trisynaptic loop. Pathways specific to CA3-CA1 synapses are framed in a blue box, pathways specific to DG-CA3 in a green box, those common to these 2 synapses in a yellow box and, finally, pathways specific to EC-DG synapses are in an orange box. Relative protein abundance for each of the 9 samples investigated by LC-MS/MS is presented as z-scores in heatmaps. A title and a heatmap is presented for each overrepresented pathway. Related pathways (i.e. CA3-CA1 pathways involved in Energy Metabolism) are framed with a dashed black line. For some pathways (i.e. Traffic of A2-containing AMPARs) we also present a heatmap with proteins that have a clear DE but did not reach statistical significance (Not Significant). In the 'Rhos Activate Wasps and Waves' gene names of members of the Arp2/3 complex are in bold.

Proteins with highest expression in CA3-CA1 synapses regulate AMPARs traffic, clathrin-mediated endocytosis, actin polymerization, Wnt signalling and glucose metabolism.

We identified the Gria2 subunit of AMPA receptors (AMPAR, Fig. 2d.1 and 3) with highest expression in CA3-CA1 synapses. Moreover, Gria3 (q = 0.067) and Gria1 (q = 0.18) presented the same trend (Gria4 was not detected). These findings suggest that CA3-CA1 synapses would have more of Gria2-containig AMPARs. Indeed, the pathway 'Traffic of Gria2 containing AMPAR' was the most overrepresented of all (fold enrichment, 38.6, Suppl. Table 4). Other proteins involved in the regulation of AMPAR traffic, such as those controlling clathrin-mediated endocytosis[44] and neuronal pentraxin 1 (Nptx1)[45], were also strongly enriched in CA3-CA1 synapses.

Although actin-related categories were found in all synaptic types (Suppl. Fig. 3f and Suppl. Table 4), CA3-CA1 synapses presented many more functional categories related to microfilaments, particularly to their polymerization. For example, all 7 members of the Arp2/3 complex, necessary for actin branching and dendritic spine structural plasticity[46], presented higher abundance in this synaptic type, albeit only three reached statistical significance (Fig 3 and Suppl. Table 2). This would be suggestive of a more refined control of spine structural dynamics in these synapses.

We also found the non-canonical Wnt signalling pathway that controls calcium levels and synaptic plasticity[47,48] overrepresented in CA3-CA1 synapses. Among the downstream effectors of this pathway, calcineurin (Ppp3ca) and the calcium-activated protein kinase C (PKC, isoenzyme Prkcg) were overexpressed in this synaptic type. Suggesting that the modulation of spine calcium dynamics via Wnt signalling might be especially relevant in these synapses.

Finally, multiple functional categories related to energy production were specifically overrepresented in CA3-CA1 synapses. Suggesting that these synapses would have higher energetic demands. These include proteins regulating the traffic of glucose transporters to the plasma membrane, five out of the 10 glycolytic enzymes and enzymes related to pyruvate metabolism or ATP synthesis.

Overexpressed proteins in DG-CA3 synapses control metabotropic and ionotropic glutamate receptors, organize neurofilaments and are involved in protein translation.

The postsynaptic metabotropic glutamate receptor Grm1 presented increased abundance in DG-CA3 synapses. This was 3.4- and 1.5-fold higher than in CA3-CA1 and EC-DG synapses, respectively. Grm1 signals through Gq protein alpha subunits, which regulate

levels of the second messengers inositol trisphosphate (IP3) and diacyl glycerol (DAG). The signalling pathways 'G alpha Q signalling events' and 'DAG and IP3 signalling' were found significantly enriched in DG-CA3 synapses. Also, Necab2 and Homer3, known to modulate metabotropic glutamate signalling[49] were found strongly overexpressed in DG-CA3 synapses.

Overexpressed proteins in DG-CA3 synapses also regulate NMDA and AMPA receptors. We found overrepresented pathways related to NMDA receptor function, including, 'Regulation of NMDA Receptor Activity' or 'Negative Regulation of NMDA Receptor Mediated Neuronal Transmission' (Suppl. Table 4). Among proteins controlling NMDARs, PTK2B might be particularly relevant, as this kinase also interacts with Grm1[50]. We also identified proteins regulating AMPAR function, including Shisa6[51], Syt17[52], Snap47[53], and Nptxr[45]. Also related to the function of both AMPA and NMDA receptors is the signalling through ERK1/2 kinases. The GO pathway 'Positive Regulation of ERK1 and ERK2 Cascade' was also found overrepresented in DG-CA3 synapses.

Interestingly, among NMDAR related proteins we identified the neurofilament light chain (Nefl), known to be involved in its trafficking[54,55]. Actually, the four proteins that form neurofilaments were found significantly overexpressed in DG-CA3 synapses. Being amongst the proteins with larger abundance differences between DG-CA3 and CA3-CA1 synapses (Suppl. Table 2). Many modulators of the Rho family of small GTPases, including GTPase activating proteins (GAPs) and, specially, guanine nucleotide exchange factors (GEFs) were also found overexpressed. This suggests that pathways regulated by these signalling molecules, mostly related to the regulation of the cytoskeleton, might be controlled in a more specific manner in this synaptic type.

Finally, we observed a very striking increase of virtually all ribosomal proteins in DG-CA3 synapses, with 21 of them reaching statistical significance (Fig. 3, Suppl. Tables 2 and 4). Moreover, several functional categories related to proteostasis were overrepresented in this synaptic type, including 'Protein Stability', or 'Unfolded Protein Binding'. Finally, Pura and Purg, involved in the transport of messenger RNA into the postsynapse[56], were also found overexpressed. To further investigate this finding, we went back to the analysis of scRNAseq done with the neurons that are engaged in the trisynaptic loop (Suppl. Table 3). In line with our proteomics findings, we observed a very strong upregulation of most ribosomal genes in the dentate gyrus (Suppl. Fig. 4a). These findings, together with the recent discovery that local translation occurs at Mossy Fibre boutons[57], indicate that proteostasis would play a particularly relevant role in this synaptic type.

Upregulated proteins in EC-DG synapses would grant them a unique extracellular matrix.

The proteome of EC-DG synapses presented several highly expressed proteoglycans, including Bcan, Ncan, Agrn and Hspg2 (Vcan and Cspg5 also presented highest expression in EC-DG, but did not reach statistical significance, Fig. 3). The synaptic location of all these proteins is well documented[2], mostly localizing to the extracellular matrix (ECM). Indeed, the GO term 'Extracellular matrix structural constituent' and the Reactome pathway 'Integrin cell surface interactions', related to the ECM, were overrepresented in EC-DG synapses. We thus observed a differential composition of the EMC, especially regarding the abundance of proteoglycans, that could specifically modulate the properties of this synaptic type. As in the previous two synaptic types, we also identified overexpressed proteins that are related to the regulation of AMPAR. These include the 'receptor-type tyrosine-protein phosphatase delta' (Ptprd)[58], the AMPAR auxiliary protein Shisa9, first described in the DG[59], and the scaffolding protein Epb41l1, known to bind to A1 subunits of the AMPAR[60,61], regulating its activity-dependant insertion into the plasma membrane[62].

Proteins with highest expression in EC-DG synapses also retrieved several pathways related to the catabolism of branched chain amino, including 'valine, leucine and isoleucine degradation' (KEGG), 'branched chain amino acid catabolism' (Reactome) or 'alpha amino acid metabolic process' (GO). One of the two metabolic pathways to synthesize glutamate requires the catabolism of these amino acids, and the product of this reaction feed into the TCA cycle. EC-DG synapses might have a preferential use of this glutamate synthesis pathway, coupling synaptic transmission with energy production.

Genes coding for proteins involved in glutamate receptor function are differentially expressed in most types of excitatory neurons.

The fact that synaptic types, formed by different neurons, all exhibit distinct expression patterns of proteins involved in the regulation of glutamate receptors prompted us to investigate whether this is a result of genetic factors. Even more so when we consider our previous finding that gene expression plays a role in synaptic proteome variability (Fig. 2g-l). To do so we investigated gene expression in excitatory neurons of the hippocampus and subiculum, using data from the Allen Brain Cell Atlas (ABCA)[34], which defines 55 types of excitatory neurons, grouped in 8 classes. We first split all genes in two groups, those coding for our reference proteome (Suppl. Table 1), which we refer to as 'synaptic genes', and the rest (non-synaptic genes). We found that 18% of synaptic genes presented expression differences between neuronal classes (Suppl. Fig. 4b, Suppl. Fig. 5a, and Suppl. Table 5) and 17% between neuronal types (Suppl. Fig. 5b-i and Suppl. Table 6). Interestingly, the frequency of DE synaptic genes was 3 times higher than in the group of non-synaptic genes

(Chi-square Test p < 0.0001, Suppl. Fig.4c). This remained significant if synaptic genes were compared to random gene sets of the same size taken from: i) all genes or ii) non-synaptic genes (Suppl. Fig.4c).

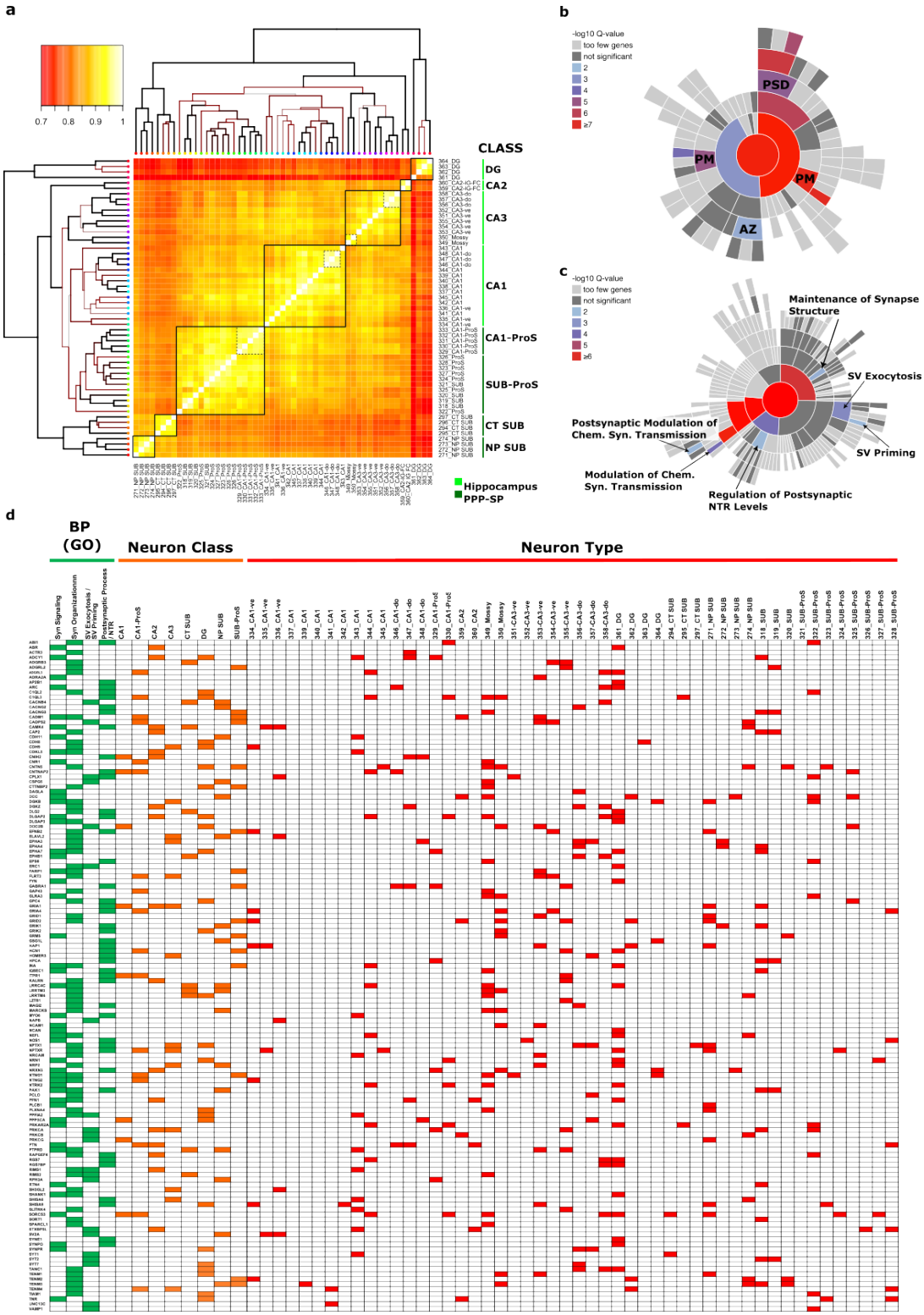


Figure 4. Differentially expressed genes mostly regulate neurotransmitter receptor function and synaptic vesicle exocytosis.

a. Clustering of the coefficients of correlation for RNA expression of up-regulated genes with a synaptic location in excitatory neuron types from the hippocampal formation.

b. Sunburst chart showing SynGO Cellular Component terms enriched among genes expressed at synapses that present increased expression in one or two types of excitatory neurons from the hippocampal formation. The background set for this analysis was the set of genes with a synaptic expression. Maximum stringency was applied for evidence filtering of SynGO annotations. PM: plasma membrane, AZ: active zone and PSD: postsynaptic density.

c. Sunburst chart showing SynGO Biological Process terms enriched among genes expressed at synapses that present increased expression in one or two types of excitatory neurons from the hippocampal formation.

d. Classes and types of excitatory neurons presenting increased expression of genes within Biological Process (GO) terms most overrepresented in the SynGO analysis.

We observed that upregulated genes were mostly present in one neuronal class, and eventually in two (Suppl. Fig. 4d), while downregulated ones appeared more repeatedly, in up to 5 classes (Suppl. Fig. 4e). The same happened in the comparison between neuronal types (Suppl. Fig. 4f), downregulated genes appeared more repeatedly. As our goal was to capture the functional categories most unique to each class or types, we only considered upregulated genes for subsequent analysis.

Next, we wanted to compare the expression patterns of upregulated synaptic genes between neuronal types. To achieve this, we computed expression correlation coefficients of each pair of neurons and performed hierarchical clustering. Surprisingly, neurons from the same class were grouped together (Figure 4a), perfectly replicating the classification obtained by the ABCA with the entire transcriptome[34]. Suggesting that synaptic genes from closely related neurons have more similar expression patterns, but also that synaptic genes have a role in the classification of hippocampal neuronal types.

To investigate common features among upregulated synaptic genes, we performed enrichment analysis of 'Cellular Component' and 'Biological Process' categories with the SynGO database. To obtain highly specific categories we used our reference proteome as a background set, and the most stringent criteria for evidence filtering. The first analysis found that these genes code for proteins residing in two main locations, the postsynaptic density (PSD) and the active zone (AZ) (Fig. 4b). The analysis of Biological Processes returned categories related to synaptic vesicle exocytosis and to the regulation glutamatergic transmission, including the regulation of neurotransmitter receptor levels (Fig. 4c). Thus, synaptic genes with more different expression patterns between neurons would be mostly related to the exocytosis of synaptic vesicles and the regulation of glutamate receptor function. Finally, we asked if the genes linked to these SynGO categories were spread across neuronal classes and types or if, instead, they were concentrated in a small number of them. We found that genes from these functional categories are widely spread across neuronal

529 classes and types (Fig. 4d), indicating that their differential regulation is a common trend
530 among them.
531
532 We also investigated the signalling pathways associated to upregulated genes from
533 individual neuronal types using pathfindR. In many instances the number of upregulated
534 genes was small (Suppl. Table 6), accordingly, pathfindR could only found enriched terms in
535 22 of the 55 neuronal types of the hippocampus and subiculum (Fig. 5a and Suppl. Table 7).
536 Nevertheless, we observed that many of the enriched pathways were again related to the
537 function of glutamate receptors (Fig. 5b). In 11 of the 22 types, upregulated genes were
538 associated with pathways related to neurotransmitter receptor function, and in 8 this term
539 was the most enriched one (Figure 5b, dark blue bars). These included terms such as
540 'ionotropic glutamate receptor activity', 'Trafficking of AMPA receptors', 'activation of NMDAR
541 and postsynaptic events' or 'extracellular ligand gated ion channel activity'. In one neuronal
542 type (CA1-343) the term 'SV exocytosis' was identified as the most enriched (Figure 5b, red
543 bars). These observations matched the findings obtained with SynGO (Fig. 4b,c), and
544 strengthen them, as they were obtained with different databases and bioinformatic tools.

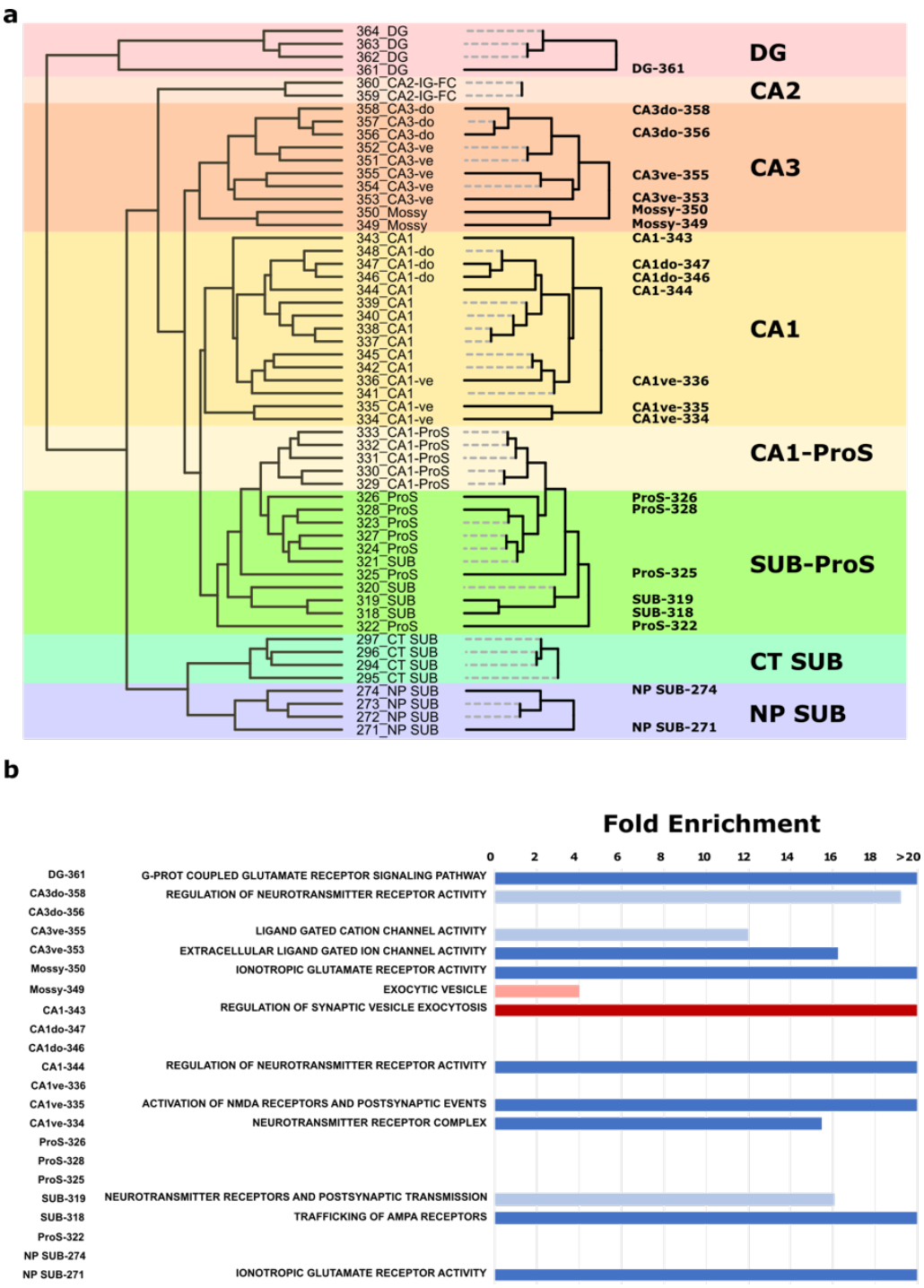


Figure 5. Hippocampal synaptic types are mostly defined by genes regulating neurotransmitter receptor function.

a. Neuron types having genes expressed at synapses that show increased expression define neuron-specific synaptic types. Dashed lines correspond to neuron types whose upregulated genes cannot be linked to significantly overexpressed term. These synapses would not present any functional difference with those of other neurons from the same class.

b. Fold enrichment of significantly enriched terms related to neurotransmitter receptor function (blue bars) or synaptic vesicle exocytosis (red bars). Dark blue or red denotes a term that is the most enriched one for that synaptic type. Light colours denote terms that are enriched but are not the most enriched. Fold enrichment corresponds to the ratio between the number of observed and expected genes related to one term.

Genes coding for proteins involved in glutamate receptor function importantly contribute to transcriptomic-based neuronal classifications.

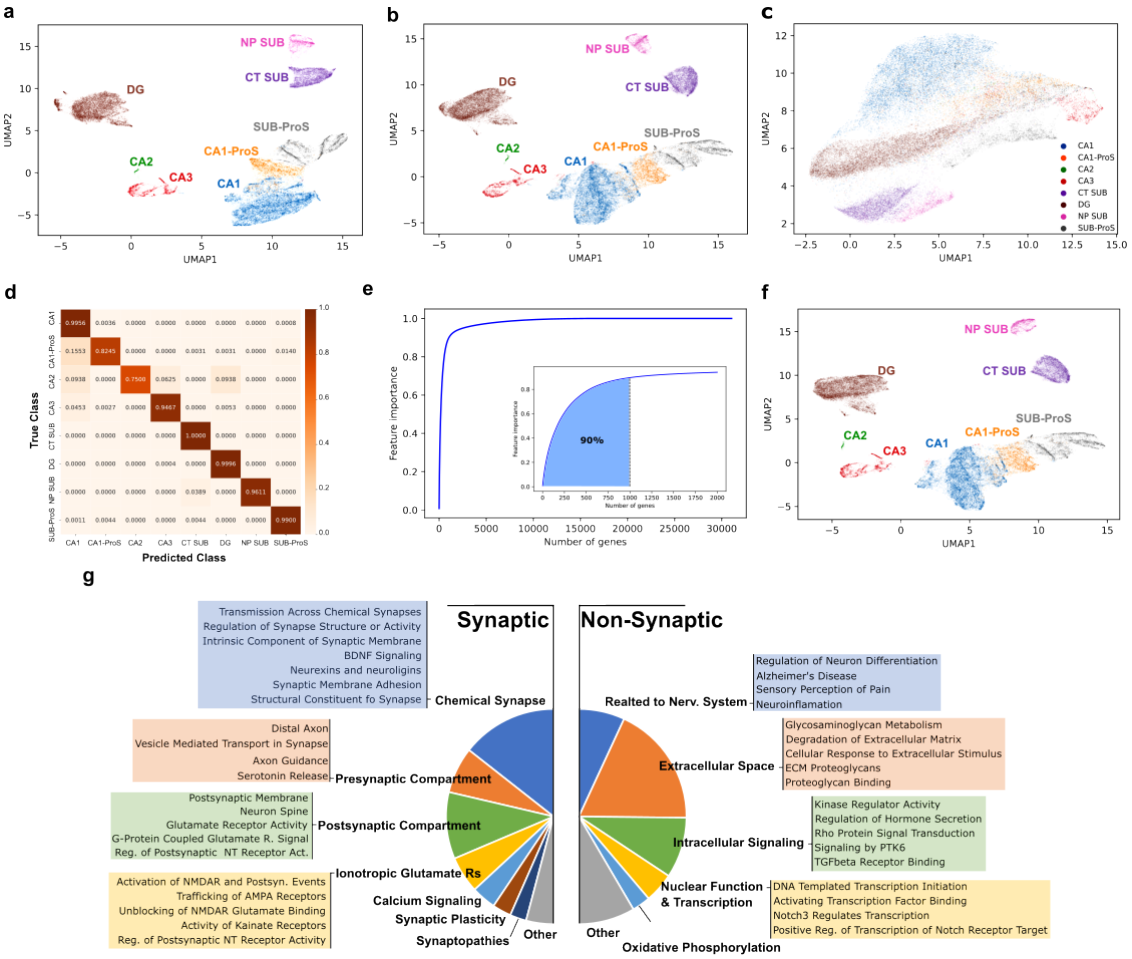
We have shown that synaptic genes generally present higher transcriptomic variation (Suppl. Fig. 4d) and that the ABCA classification of hippocampal neurons[34] can be replicated only using upregulated synaptic genes (Fig. 4a). Suggesting that synaptic genes play an important role in determining hippocampal neuronal types. To investigate this possibility, we referred again to the ABCA database. We first confirmed that we could replicate the ABCA classification with the entire transcriptome, as classes of excitatory neurons clearly segregated in nonlinear dimensionality reduction maps (U-Map) (Fig. 6a). Noticeably, the U-map generated with synaptic genes (Fig. 6b) was highly similar to that produced with all genes. Instead, U-Maps from non-synaptic genes (Fig. 6c and Suppl. Fig. Suppl_6a) had very different topologies, with a much higher overlap between neurons from different classes. Indicating that synaptic genes importantly contribute to the classification of hippocampal excitatory neurons, as it has been recently shown for cortical neurons[63]. To further validate this observation, we asked how many of the genes that contributed most to the classification were synaptic. To identify genes with a large contribution to the classification we used the Random Forest method, a supervised machine learning approach for data classification[64], that determines the importance of each variable (here gene expression data) in a classification problem.

After the training phase, the algorithm could predict the class of a given neuron with high accuracy (total accuracy for the train set 0.9893 - total accuracy for the test set 0.9014), indicating that the algorithm effectively replicated the classification, and that the computed weight of each gene to the classification was reliable. Indeed, the predictive power of the algorithm was above 95% for 6 of the 8 neuronal classes (Fig. 6d). A small number of genes did drive the overall classification. The added weight of the 1000 genes most contributing to the classification accounted for 90% of the information carried by the whole transcriptome (Fig. 6e and Suppl. Table 8). Importantly, over 50% of this top 1000 genes were synaptic (Suppl. Fig.6b1), a 4x overrepresentation that was highly statistically significant (Chi-square test, $p < 1e\text{-}23$). Using the synaptic genes in the top 1000 was sufficient to replicate the U-map generated with the entire transcriptome (Fig. 6f). Furthermore, the accuracy of the Random Forest prediction was better when using all synaptic genes as opposed to the entire transcriptome and best when using the synaptic genes found in the top 1000 list (Suppl. Fig.6c).

Using the Chi-square Stat value, we found that genes expressed at synapses were more over-represented in genes driving the classification than genes enriched in the PSD[65], in the MASC complex[66] or in other functional categories enriched in the top 1000 genes contributing

to the classification. (Suppl. Fig.6d). Random Forest performance was also good in classifying neurons into types, although less accurate (total accuracy of the train set 0.8559 and total accuracy of the test set 0.7653, Suppl. Fig. 6e). The list of the top 1000 genes most relevant to the classification of types also carried over 90% of the weight, and included over 500 synaptic genes (Chi-square test, $p < 1e-10$, Suppl. Fig6b.2 and Suppl. Table 8).

pathfindR analysis of synaptic genes in the top 1000 most contributing to the classification of neuronal classes revealed synaptic functions or locations related to both pre and postsynaptic compartments (Fig. 6g,h and Suppl. Table 8). Yet, those terms with highest fold enrichment were mostly related to the function and organization of glutamate receptors (Fig. 6h). Curiously, non-synaptic genes of the top 1000 genes were also associated with some functions of the nervous system (i.e. Neuron differentiation or Neuroinflammation), among others.



h

#	Term Description	Database	Fold_Enrichment	Adjusted p-value
1	IONOTROPIC ACTIVITY OF KAINATE RECEPTORS	Reactome	22,9	7,42E-12
2	G-PROTEIN COUPLED GLUTAMATE RECEPTOR SIGNALING PATHWAY	GO-Biological Process	22,6	1,36E-05
3	REGULATION OF SHORT TERM NEURONAL SYNAPTIC PLASTICITY	GO-Biological Process	20,8	1,14E-02
4	GLUTAMATE RECEPTOR ACTIVITY	GO-Molecular Function	20,6	6,23E-18
5	UNBLOCKING OF NMDA RECEPTORS GLUTAMATE BINDING AND ACTIVATION	Reactome	20,2	2,10E-25
6	NEURON CELL CELL ADHESION	GO-Biological Process	20,1	1,87E-06
7	GPCRS CLASS C METABOTROPIC GLUTAMATE PHEROMONE	Wikipathways	19,1	3,09E-10
8	SEROTONIN NEUROTRANSMITTER RELEASE CYCLE	Reactome	18,2	8,97E-19
9	DISRUPTION OF POSTSYNAPTIC SIGNALLING BY CNV	Wikipathways	17,4	3,56E-25
10	TRAFFICKING OF AMPA RECEPTORS	Reactome	16,9	8,75E-32

Figure 6. Expression differences in genes encoding synaptic proteins strongly determine the classification of excitatory neurons.

a. UMAP graph generated with single-cell RNA abundance data obtained from excitatory neurons in the 8 classes identified in the hippocampal formation. Abundance of all genes in the genome was considered for the construction of this graph. ProS, prosubiculum; SUB, subiculum, NP SUB; near-projecting neurons from the subiculum and CT SUB; corticothalamic neurons from the subiculum.

b. UMAP generated as in (a), although in this occasion only genes coding for synaptic proteins were considered.

c. UMAP generated as in (a), using a random set of genes not expressed at synapses, with the same number of genes as in the synaptic dataset in (b).

d. Confusion or error matrix generated by the Random Forest algorithm, showing the success rates in assigning a class to each neuron. Colour legend correspond with the accuracy of the prediction, 1 being perfect accuracy.

e. Cumulative importance of the expression level of each gene in the genome for the classification of excitatory neurons into classes. Inset, cumulative Importance of the top 2000 genes with the highest importance to the classification. Note that the top 1000 contributing genes provide 90% of the information necessary to construct the classification.

f. UMAP generated as in (a) but using only the 520 synaptic genes found among the Top 1000 genes contributing to the classification.

g. Main signalling pathways and biological functions found among genes encoding for synaptic and non-synaptic proteins of the top 1000 that most contribute to the classification of excitatory neurons into classes.

h. Top 10 signalling pathways with the largest fold enrichment. In yellow those relative to the function of ionotropic or metabotropic glutamate receptors. In green those relevant to presynaptic function.

Discussion

Electrophysiological studies show that different synaptic types have unique functional properties[67,68]. Yet the molecular basis driving these differences are poorly understood. Investigating synaptic types at the proteomic level has been extremely challenging, as they are confined to microscopic brain regions. To overcome this limitation, we have developed a procedure to obtain microscopic brain samples containing individual synaptic types, and to extract synaptic proteins from them in sufficient quantity for subsequent high-throughput proteomics. This method has several advantages, first it provides a great level of anatomical resolution, since the exact location of collected samples is known. Moreover, it delivers a wide coverage of the synaptic proteome, identifying proteins from most subsynaptic compartments. Lastly, it can be used in any species, including humans, as it does not require prior genomic manipulations. With this approach we have extensively profiled the proteome of the synaptic types that constitute the trisynaptic circuit of the hippocampus.

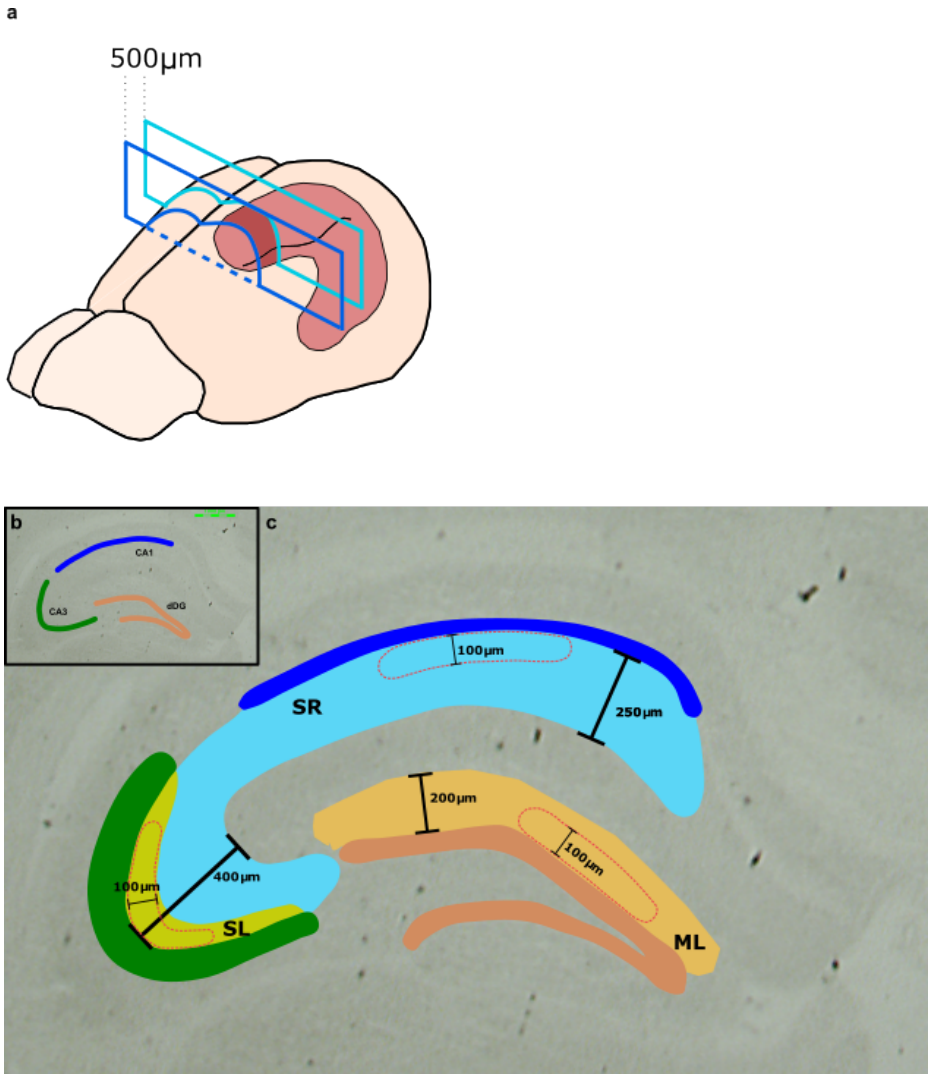
An important conclusion of our proteomics data is that essentially the same proteins are present in the three synaptic types investigated. This observation is relevant, as it implies that functional diversity among them arises from changes in the abundance of shared components. These would result in specific molecular processes being differentially favoured between synaptic types. For example, it is well-known that CA3-CA1 synapses require NMDARs activation for LTP expression but DG-CA3 synapses don't. Several synaptic types express forms of NMDAR-independent LTP across the brain, and class I metabotropic glutamate receptors (Grms) are involved in some of them[69,70]. Indeed, the role of Grm1/5 in NMDAR-independent LTP at DG-CA3 synapses has been addressed by a few studies, albeit these returned contradictory results[17]. Our data provides strong support for a role of Grm1 in NMDAR-independent LTP in DG-CA3 synapses, as this receptor and several of its downstream signalling molecules are highly expressed in them. Thus, while all these molecules are present in both synaptic types, the increased abundance of Grm1 and its downstream signalling proteins in DG-CA3 synapses would provide them with the ability to express an NMDAR-independent form of LTP, finetuning the functional properties of this particular synaptic type.

Proteins differentially expressed between synaptic types were implicated in many signalling pathways and biological processes related to synaptic biology. Remarkably, the vast majority of these were exclusively found in one synaptic type. Suggesting that they could contribute specific functions to different synaptic types. CA3-CA1 synapses exhibited several overrepresented pathways directly related with AMPAR traffic, but also to clathrin mediated endocytosis, the primary mechanism by which AMPARs are removed from the synapse[44]. These synapses also displayed many functional categories related to actin polymerization

673 and branching, key processes in spine structural plasticity. The non-canonical Wnt/Ca^{2+}
674 pathway, which regulates calcium release from internal stores[47], was also overrepresented
675 in this synaptic type. And so were numerous metabolic pathways related to energy production
676 in CA3-CA1 synapses, suggesting they might have increased energetic demands. Instead,
677 DG-CA3 synapses were characterised by signalling pathways downstream of class I
678 metabotropic glutamate receptors. They also exhibited a sticking increase in ribosomal
679 proteins, likely due to an elevated number of presynaptic ribosomes, as protein translation at
680 mossy fibre boutons would regulate synaptic plasticity[57]. They also presented increased
681 levels of proteins that positively regulate ERK1/2 signalling, a pathway linking ionotropic
682 glutamate receptors with protein translation. In line with previous findings, showing that
683 mossy fibre boutons have the highest level of ERK1/2 activation in the hippocampus[71].
684 Furthermore, DG-CA3 synapses presented increased abundance of all 4 proteins organizing
685 intermediate neurofilaments. These proteins have been confidently identified in synapses[54],
686 being involved in synaptic transmission and plasticity[54]. Our data indicates that CA3-CA1 and
687 DG-CA3 synapses would have specific requirements regarding their cytoskeletal function.
688 Structural plasticity at the level of dendritic spines has been investigated with two-photon
689 microscopy, albeit in cortical neurons[72]. However, these studies show considerable
690 differences between neurons, differences which might arise from different cytoskeletal
691 compositions. Finally, EC-DG synapses were strongly characterised by a unique ECM, with
692 increased levels of several proteoglycans and other constituents of the ECM. The synaptic
693 localization of proteoglycans is also well documented[2], contributing to AMPAR traffic[73,74] and
694 synaptic transmission[75]. Indeed, the ECM as a whole is known to restrict AMPAR mobility[76].
695
696 Overall, our proteomic findings provide support for considerable molecular diversity among
697 the synaptic types of the trisynaptic loop. Impacting multiple domains of synaptic biology,
698 including the traffic and synaptic stability of AMPARs, spine structural plasticity, signalling
699 through metabotropic receptors, control of calcium levels, local protein translation or
700 regulation of the energetic metabolism, among others. However, it is also important to
701 mention that in all synaptic types we found among differentially expressed proteins molecules
702 that regulate the function of glutamate receptors. Being that these synaptic types are formed
703 by 4 different neurons, and that we have shown that gene expression contributes to synaptic
704 proteome diversity, we decided to investigate if gene expression mechanisms contributed to
705 the common regulation of these proteins.
706
707 Having first identified the synaptic genes differentially expressed between neuronal types we
708 next looked for the functional categories most related to them. A first analysis of all these
709 genes together identified that they mostly localize to two subsynaptic locations, the active
710 zone and the postsynaptic density. Being involved in synaptic vesicle (SV) exocytosis, and

711 the postsynaptic regulation of chemical synaptic transmission, especially the regulation of
712 neurotransmitter receptor levels at the synapse. Importantly, genes involved in these
713 processes were differentially expressed in most neuronal types, with each type
714 overexpressing a subset of them. Therefore, the differential regulation of these proteins is a
715 common trend among excitatory neurons in the hippocampus and subiculum. In a second
716 analysis we investigated the signalling pathways overrepresented in independent neuronal
717 types. These analyses also retrieved many pathways related to glutamate receptor function,
718 actually, these were the most enriched ones for many neuronal types. Pathways related to
719 SV exocytosis weakly overrepresented in the analysis of individual neuronal types.
720
721 An orthogonal mathematical approach based on machine learning corroborated the
722 differential expression of genes related to glutamatergic function between neuronal types.
723 This approach was employed to identify the genes that contribute the most to
724 transcriptomics-based neuronal classifications. This analysis demonstrated that genes
725 involved in glutamatergic function were key to these classifications, as they presented unique
726 expression patterns across neuronal types. Thus, we propose that the differential regulation
727 of these genes is a main feature of gene expression programs between neuronal types. This
728 finding aligns with the recent observation that, in general, synaptic genes contribute to the
729 classification of cortical neurons[63].
730
731 In the present study, we introduce a novel procedure to isolate individual synaptic types and
732 analyse their proteome. With this method we have been able to identify major molecular
733 differences between the synaptic types that comprise the trisynaptic circuit. This is an
734 important resource to advance in our understanding of the molecular mechanisms controlling
735 their diverse electrophysiological properties. More importantly, our combined investigation of
736 proteomic and transcriptomic datasets indicates that glutamate receptors and proteins
737 directly controlling their function, are common drivers of synaptic proteome variability across
738 synaptic types. Thus having key contributions to the properties of different synaptic types. It
739 is interesting to note that neuron-specific transcriptional mechanisms would contribute to the
740 unique expression levels of these proteins.

Supplementary Figures and Legends

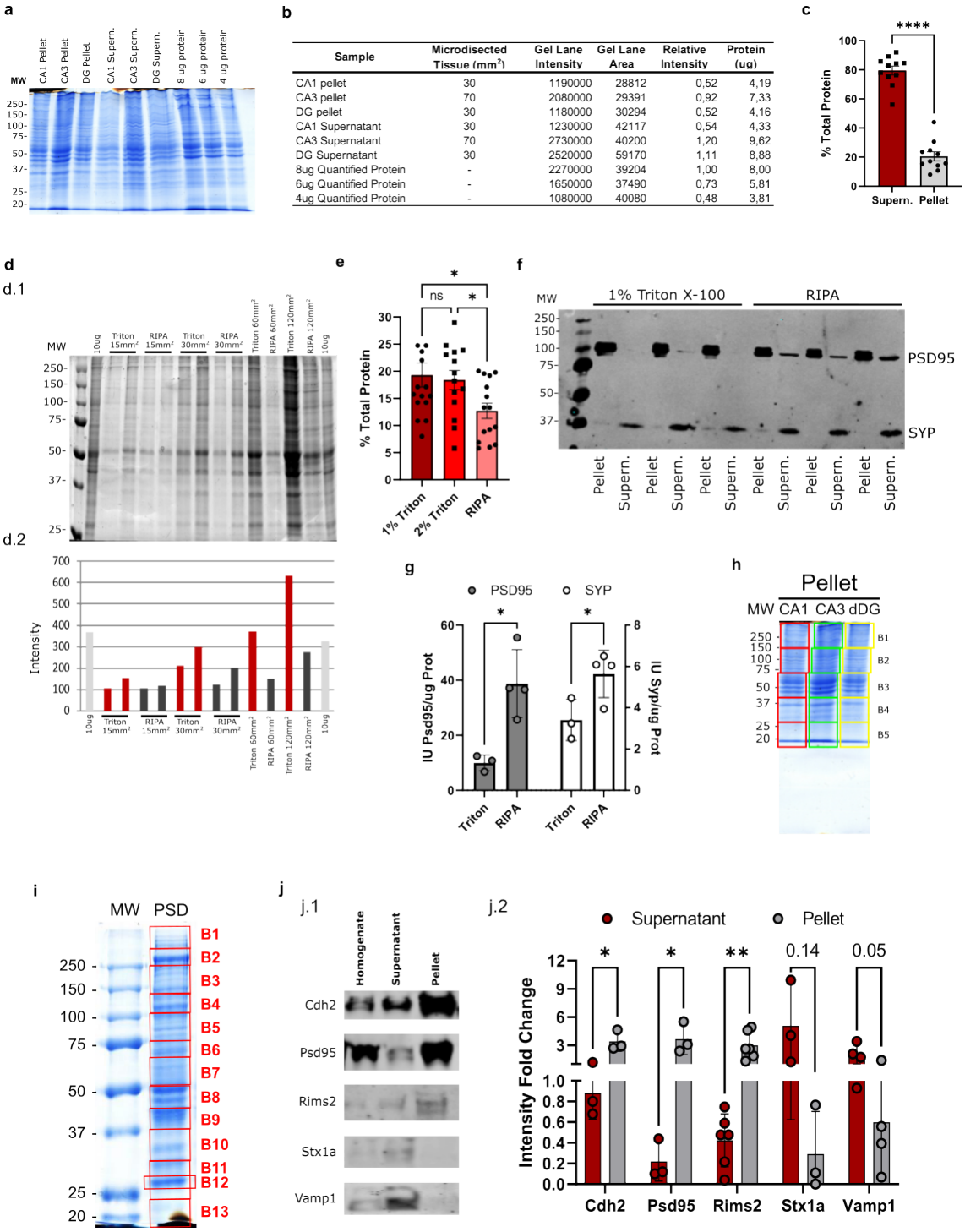


Supplementary Figure 1. Strategy used to microdissect the hippocampal layers containing the synapses of the trisynaptic loop.

a. Drawing of mouse brain showing the localization of the hippocampus, in red. The portion of the dorsal hippocampus that was analysed in this work is shown in dark red. This covered approximately 500 µm in the longitudinal axis of the brain.

b. Brightfield image showing the hippocampus in a coronal section of the dorsal mouse brain. The 3 subfields, CA1, CA3 and dDG, investigated in this study are indicated. CA1 and CA3 pyramidal layers and dDG granular layer are shown differently coloured. Scale bar 1000µm.

c. Anatomical localization and dimensions, particularly width, of the different hippocampal layers from which we collected microdissected neuropile. Pyramidal and granular layers coloured as in (a). Shapes delimited by red dashed lines represent examples of microdissected neuropile fragments, where their width is also shown. Collected fragments in all subfields had 100µm in width, approximately. Neuropile fragments were collected from the following layers: i) Stratum Radiatum (SR, in blue) in the CA1 subfield, ii) from the Stratum Lucidum (SL, in pale green) in the CA3 subfield and iii) from the Molecular Layer (ML) at the dorsal Dentate Gyrus.



Sample	Microdisected Tissue (mm²)	Gel Lane Intensity	Gel Lane Area	Relative Intensity	Protein (ug)
CA1 pellet	30	1190000	28812	0,52	4,19
CA3 pellet	70	2080000	29391	0,92	7,33
DG pellet	30	1180000	30294	0,52	4,16
CA1 Supernatant	30	1230000	42117	0,54	4,33
CA3 Supernatant	70	2730000	40200	1,20	9,62
DG Supernatant	30	2520000	59170	1,11	8,88
8ug Quantified Protein	-	2270000	39204	1,00	8,00
6ug Quantified Protein	-	1650000	37490	0,73	5,81
4ug Quantified Protein	-	1080000	40080	0,48	3,81

Supplementary Figure 2. Optimization and validation of the biochemical procedure used to obtain preparations enriched in synaptic structures and proteins.

a. Protein gel electrophoresis stained with Coomassie Brilliant Blue with 1% Triton X-100 insoluble (Pellet) and soluble (Supern.) fractions obtained from microdissected neuropile of the three hippocampal subfields. 8, 6 and 4 μg of precisely quantified protein from hippocampal synaptic fractions isolated by density gradient ultracentrifugation were used to estimate protein abundance of fractions derived from microdissected neuropile.

b. Quantification of signal intensity of gel lanes in (a) and estimation of protein abundance in biochemical fractions from microdissected neuropile.

c. Percentage of total protein from microdissected neuropile recovered in 1% Triton X-100 soluble (Supern.) and insoluble (Pellet) fractions. Statistics, unpaired T-Test, **** < 0.0001.

d. d.1. Silver-stained protein gel with biochemical fractions insoluble to 1% Triton X-100 or RIPA buffer obtained from increasing areas (mm^2) of microdissected neuropile. 10μg of quantified protein from hippocampal synaptic fractions isolated by density gradient ultracentrifugation were added to the gel for reference. d.2. Bar chart with signal intensity from gel lanes in d.1.

e. Percentage of protein recovered in the pellet fractions of microdissected tissue treated with a buffer containing 1% Triton X-100 (dark red column), 2% Triton X-100 (red column) or with a RIPA buffer (light red column). Statistics, One-way ANOVA and Fisher's LSD post-hoc test, * $p < 0.05$.

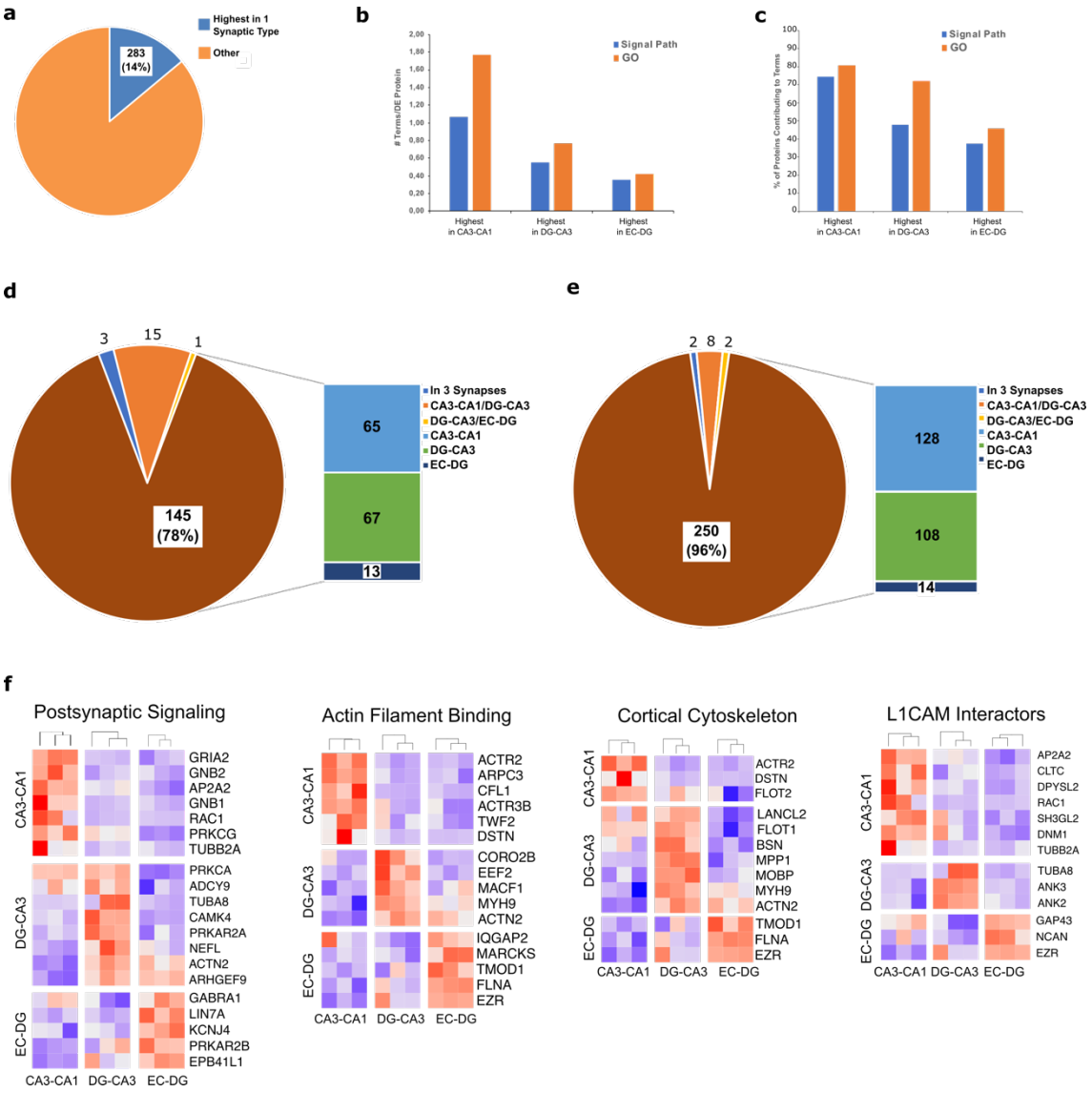
f. Immunoblot of Triton and RIPA insoluble (Pellet) and soluble (Supern.) fractions obtained from hippocampal microdissected neuropile. Proteins investigated are Psd95, mostly insoluble to triton and Synaptophysin (Syp) a synaptic vesicle protein, mostly soluble to triton.

g. Bar plot of Psd95 (grey bars) and Synaptophysin (Syp, white bars) abundance as determined by immunoblot of 1% Triton X-100 soluble fractions from microdissected hippocampal neuropile. Statistics, unpaired T-test, * $p < 0.05$.

h. Protein gel electrophoresis of 1% Triton X-100 insoluble pellets stained with Coomassie Brilliant Blue. 5 gel bands (B1-B5) were collected and processed independently in the proteomics workflow.

i. Protein gel electrophoresis of synaptic preparations isolated by density gradient ultracentrifugation from mouse hippocampi stained with Coomassie Brilliant Blue. 13 bands (B1-B13) were collected and processed independently in the proteomics workflow to generate a synaptic reference proteome.

j. j.1. Immunoblots of a transsynaptic protein involved in cell adhesion (Cdh2, cadherin 2), a postsynaptic scaffolding molecule (Psd95), and three presynaptic proteins located at: i) active zone (Rims2), SNARE complex (Stx1a) and synaptic vesicles (Vamp1). Samples analysed are whole extract (Homogenate) and triton soluble (Supernatant) and insoluble (Pellet) fractions from microdissected hippocampal neuropile. j.2. Abundance of proteins in m.1. relative to their abundance in the homogenate fraction. Statistically significant difference in protein abundance between the supernatant and pellet fractions is indicated. Statsitcs, unpaired T-test, * $p < 0.05$, ** $p < 0.01$.



Supplementary Figure 3. Most Signalling Pathways and Gene Ontology (GO) terms found in proteins differentially expressed between synaptic types are synapse-specific.

a. Proportion of proteins identified by proteomics with a statistically significant highest expression in one of the three synaptic types investigated.

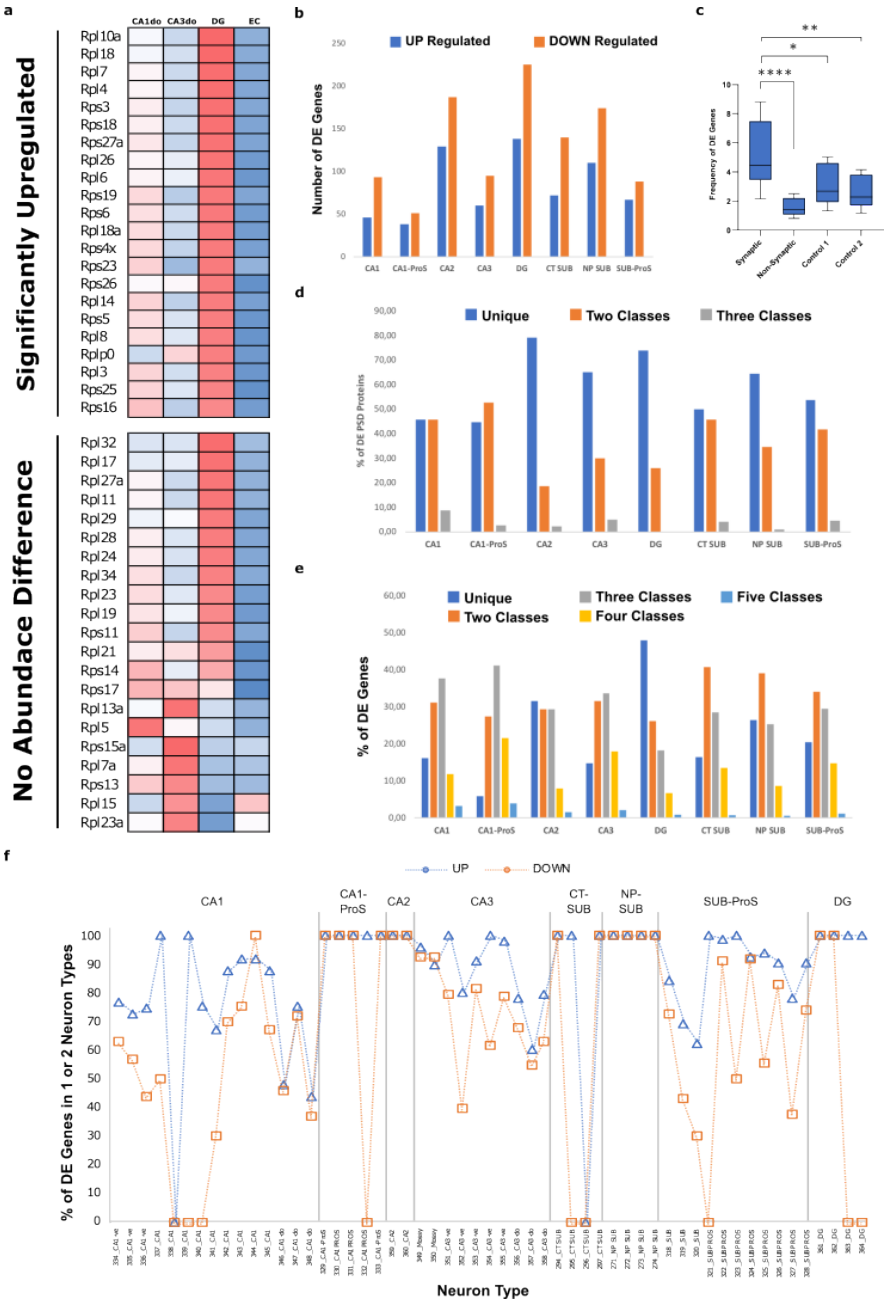
b. Ratio of significantly overrepresented terms per differentially expressed (DE) protein in each sample investigated. Signalling pathways (blue bars) investigated are from the databases Reactome, KEGG and Wikipathways. Gene ontology (GO, orange bars) terms investigated belong to the domains Cellular Component, Biological Process and Molecular Function.

c. Percentage of differentially expressed proteins contributing to signalling pathways (blue bars) and GO terms (orange bars).

d. Number of signalling pathways significantly overrepresented in 3, 2 or 1 hippocampal layers.

e. Number of significantly overrepresented GO terms among proteins with highest expression in 3, 2 or 1 hippocampal layer.

f. Heatmaps showing relative protein abundance in the 9 samples analysed by proteomics, three biological replicates per hippocampal layer, analysed in our proteomics workflow. High abundance shown in red and low in blue. Abundance of proteins in four pathways/terms found significantly enriched in all three hippocampal regions is presented.



Supplementary Figure 4. Significantly up-regulated genes are more specific to neuronal classes and types than down-regulated ones.

a. Heatmap presenting normalized (z-score) mean RNA abundance of genes coding for ribosomal proteins from the 4 classes of excitatory neurons that constitute the trisynaptic loop in the dorsal (do) hippocampus (CA1do, CA3do, dentate gyrus -DG- and entorhinal cortex -EC-). All ribosomal gens showing statistically significant RNA expression differences were upregulated in the DG (top panel). Many genes for which expression differences did not reach statistical significance also display a tendency for increased expression in the DG (bottom panel). RNA expression data obtained from the Allen Brain Cell Atlas (REFF). Statistical analysis of RNA expression differences between neuronal classes was performed with the Seurat R package and the Wilcoxon Rank Sum test. Abundance scale, 2 (dark red) to -2 (dark blue).

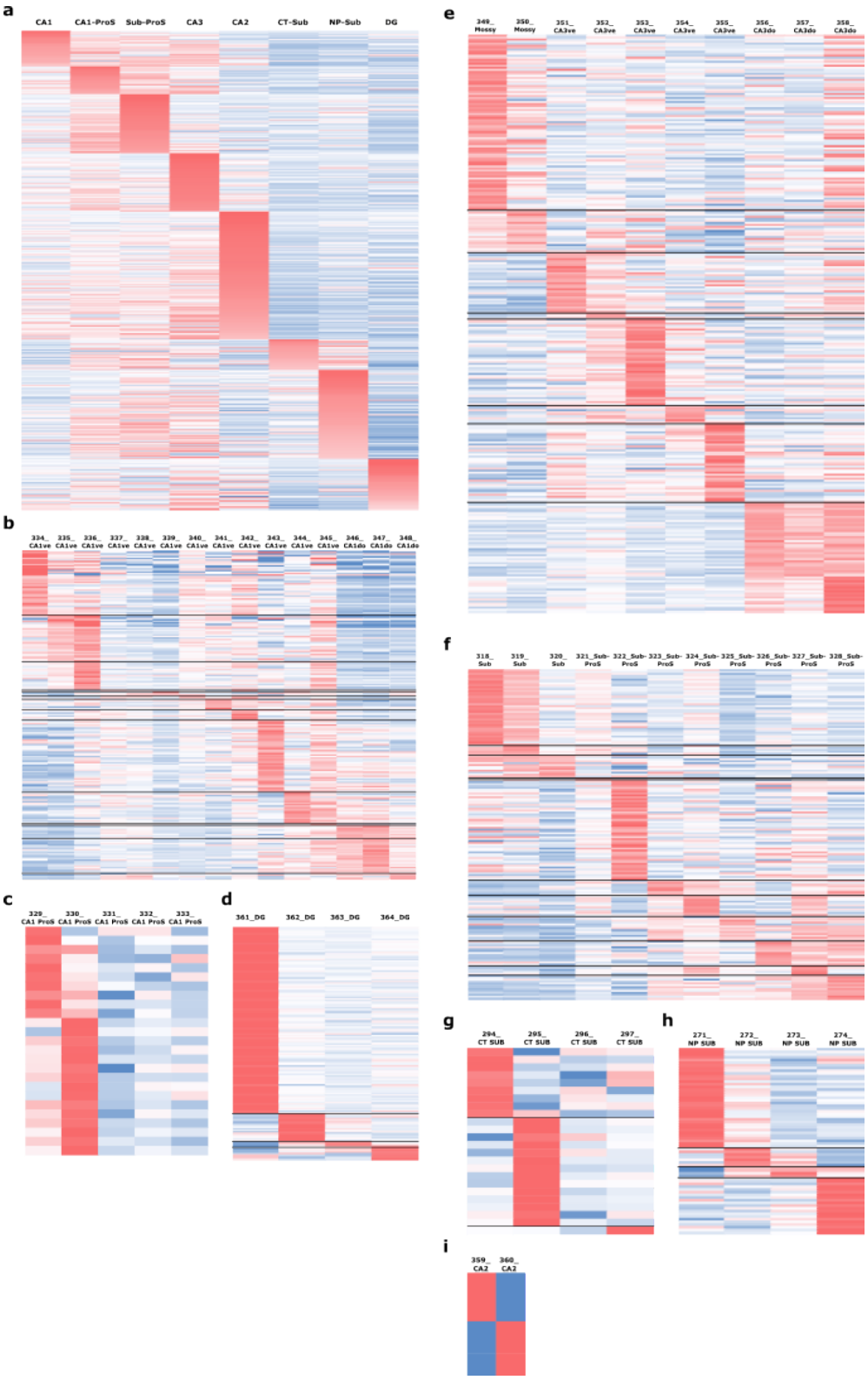
b. Number of genes expressed at synapses found significantly up- (blue bars) or down- regulated (orange bars) in classes of excitatory neurons from the hippocampal formation.

c. Frequency of differentially expressed genes among different gene sets. Including genes expressed at synapses (synaptic), genes not expressed at synapses (Non-synaptic), a random set of all genes of the same size of the synaptic set (Control 1) and a random set of non-synaptic genes of the same size of the synaptic set (Control 2). Statistics, Chi square Test, **** $p < 0.0001$, ** $p < 0.01$ and * $p < 0.05$.

d. Percentage of gens localized to synapses that are upregulated in one (blue), two (orange), or three (grey) classes of excitatory neurons.

e. Percentage of gens localized to synapses that are downregulated in one (blue), two (orange), three (grey), four (yellow) or five (light blue) classes of excitatory neurons.

f. Percentage of genes expressed at synapses being up-regulated (blue line) or down-regulated (orange line) in 1 or 2 excitatory neuron types from the hippocampal formation.



Supplementary Figure 5. Computational strategy for the identification of genes encoding synaptic proteins having increased expression in excitatory neuronal classes and types.

a. Heatmap showing relative RNA abundance data across all excitatory neuronal classes for genes found upregulated in each class.

b. Heatmap showing relative RNA abundance data across types of CA1 excitatory neurons for genes found upregulated in each type.

c. Heatmap showing relative RNA abundance data across types of CA1-ProS excitatory neurons for genes found upregulated in each type.

d. Heatmap showing relative RNA abundance data across types of dentate gyrus (DG) excitatory neurons for genes found upregulated in each type.

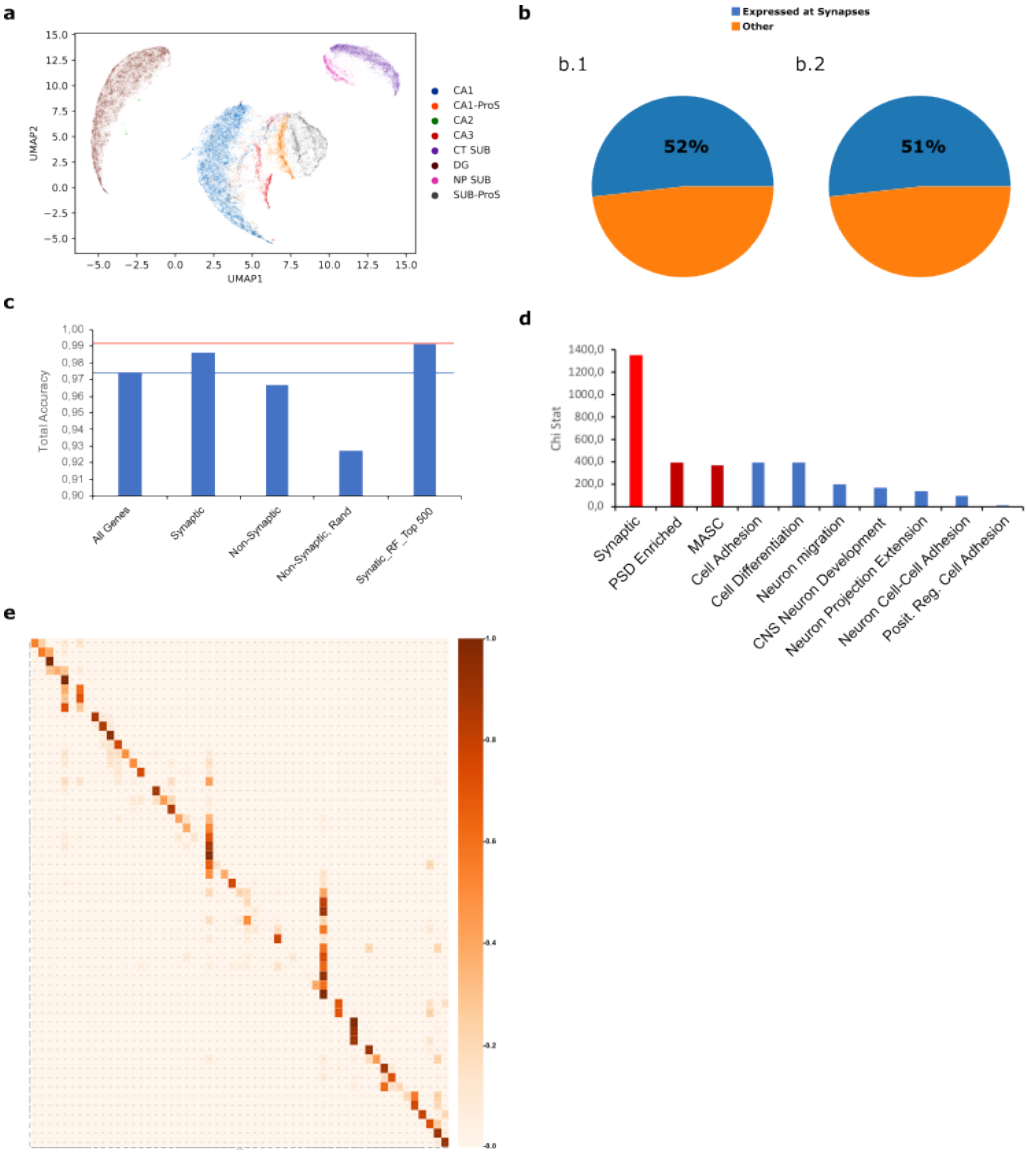
e. Heatmap showing relative RNA abundance data across types of CA3 excitatory neurons for genes found upregulated in each type.

f. Heatmap showing relative RNA abundance data across types of SUB-ProS excitatory neurons for genes found upregulated in each type.

g. Heatmap showing relative RNA abundance data across types of CT-SUB excitatory neurons for genes found upregulated in each type.

h. Heatmap showing relative RNA abundance data across types of NP-SUB excitatory neurons for genes found upregulated in each type.

i. Heatmap showing relative RNA abundance data across the two types of CA2 excitatory neurons for genes found upregulated in each type.



Supplementary Figure 6. Compared analysis of the importance of different gene sets for the classification of excitatory neurons into classes and types shows genes encoding synaptic proteins are highly relevant.

a. UMAP graph generated with single-cell RNA abundance data obtained from excitatory neurons in the 8 classes identified in the hippocampal formation This graph was generated with expression data from genes not expressed at synapses.

b. b.1. Proportion of genes with synaptic (blue) or non-synaptic (orange) localization among the top 1000 genes contributing to the classification of neuronal classes. b.2. Proportion of genes with synaptic (blue) or non-synaptic (orange) localization among the top 1000 genes contributing to the classifications of neuronal types.

c. Bar plot showing total accuracy in neuronal class prediction determined by the Random Forest machine learning method using different gene sets to train the algorithm: i) all gens in the dataset, ii) genes with a synaptic localization, iii) all genes with a non-synaptic localization, iv) a set of randomly selected genes not found at the synapse and of the same size of the set of genes localized at synapses and v) the set of 520 genes expressed at synapses among the 1000 genes mostly contributing to neuronal classification.

d. Bar plat with the Chi square statistic obtained from Chi square tests of overrepresentation of different gene sets among the 1000 genes mostly contributing to neuronal classification.

e. Confusion or error matrix generated by the Random Forest algorithm, showing the success rates in assigning a neuronal type to each neuron. Colour legend correspond with the accuracy of the prediction, 1 being maximum accuracy.

Supplementary Tables Legends

Supplementary Table 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 a PSDII fraction published by Distler et al.

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

Sheet#1: Protein abundance data generated by MSqROB 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 Table 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 stablished 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 form the ABCA for this analysis. Log2 fold changes and p-values are indicated.

Supplementary Table 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, KEEG 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).

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).

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 Table 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.

Sheet#3: List of genes significantly up-regulated in one or two classes.

Supplementary Table 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.

Sheets#3, 5, 7, 9,11, 13, 15 and 17: List of genes significantly upregulated in one or two neuronal types within each class.

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

Supplementary Table 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.

Sheet#2: List of representative GO terms identified by pathfindR for synaptic genes upregulated in different neuronal types.

Supplementary Table 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 Pathfinder from the synaptic genes among the top 1000 most contributing to the classification of excitatory neurons.

Sheet#4: Representative terms (Signalling Pathways and GO terms) identified by Pathfinder from the non-synaptic genes among the top 1000 most contributing to the classification of excitatory neurons.

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

Supplementary Video. Manual dissection of hippocampal subfields.

References

1. Li, K. W., Gonzalez-Lozano, M. A., Koopmans, F. & Smit, A. B. Recent Developments in Data Independent Acquisition (DIA) Mass Spectrometry: Application of Quantitative Analysis of the Brain Proteome. *Front. Mol. Neurosci.* **13**, 564446 (2020).

2. Koopmans, F. *et al.* SynGO: An Evidence-Based, Expert-Curated Knowledge Base for the Synapse. *Neuron* **103**, 217-234.e4 (2019).

3. Bayés, A. & Grant, S. G. N. Neuroproteomics: understanding the molecular organization and complexity of the brain. *Nature Reviews Neuroscience* **10**, 635–646 (2009).

4. O'Rourke, N. A., Weiler, N. C., Micheva, K. D. & Smith, S. J. Deep molecular diversity of mammalian synapses: why it matters and how to measure it. *Nature Reviews Neuroscience* (2012) doi:10.1038/nrn3170.

5. Dieterich, D. C. & Kreutz, M. R. Proteomics of the Synapse--A Quantitative Approach to Neuronal Plasticity. *Molecular & cellular proteomics : MCP* **15**, 368–381 (2015).

6. Bayés, A. *et al.* Comparative study of human and mouse postsynaptic proteomes finds high compositional conservation and abundance differences for key synaptic proteins. *PLoS ONE* **7**, e46683 (2012).

7. Sorokina, O. *et al.* A unified resource and configurable model of the synapse proteome and its role in disease. *Sci Rep-uk* **11**, 9967 (2021).

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

9. Bayés, A. *et al.* Characterization of the proteome, diseases and evolution of the human postsynaptic density. *Nature neuroscience* **14**, 19–21 (2011).

10. Roy, M. *et al.* Proteomic analysis of postsynaptic proteins in regions of the human neocortex. *Nature neuroscience* **21**, 130–138 (2018).

11. Trinidad, J. C., Thalhammer, A., Burlingame, A. L. & Schoepfer, R. Activity-dependent protein dynamics define interconnected cores of co-regulated postsynaptic proteins. *Molecular & cellular proteomics : MCP* (2012) doi:10.1074/mcp.m112.019976.

12. Malenka, R. C. & Bear, M. F. LTP and LTD: an embarrassment of riches. *Neuron* **44**, 5–21 (2004).

13. Zhu, F. *et al.* Architecture of the Mouse Brain Synaptome. *Neuron* **99**, 781-799.e10 (2018).

14. Cizeron, M. *et al.* A brain-wide atlas of synapses across the mouse lifespan. *Science* (2020) doi:10.1126/science.aba3163.

15. Huganir, R. L. & Nicoll, R. A. AMPARs and Synaptic Plasticity: The Last 25 Years. *Neuron* **80**, 704–717 (2013).

16. Citri, A. & Malenka, R. C. Synaptic plasticity: multiple forms, functions, and mechanisms. *Neuropsychopharmacology* **33**, 18–41 (2008).

17. Nicoll, R. A. & Schmitz, D. Synaptic plasticity at hippocampal mossy fibre synapses. *Nat Rev Neurosci* **6**, 863–876 (2005).

18. Biesemann, C. *et al.* Proteomic screening of glutamatergic mouse brain synaptosomes isolated by fluorescence activated sorting. *The EMBO Journal* **33**, 157–170 (2014).

19. Oostrum, M. van *et al.* The proteomic landscape of synaptic diversity across brain regions and cell types. *bioRxiv* 2023.01.27.525780 (2023) doi:10.1101/2023.01.27.525780.

20. Apóstolo, N. *et al.* Synapse type-specific proteomic dissection identifies IgSF8 as a hippocampal CA3 microcircuit organizer. *Nat Commun* **11**, 5171 (2020).

21. Luquet, E., Biesemann, C., Munier, A. & Herzog, E. Purification of Synaptosome Populations Using Fluorescence-Activated Synaptosome Sorting. in vol. 1538 121–134 (Springer New York, 2016).

22. Paget-Blanc, V. *et al.* A synaptomic analysis reveals dopamine hub synapses in the mouse striatum. *Nat. Commun.* **13**, 3102 (2022).

23. Uezu, A. *et al.* Identification of an elaborate complex mediating postsynaptic inhibition. *Science* **353**, 1123–1129 (2016).

24. Takano, T. & Soderling, S. H. Tripartite synaptomics: Cell-surface proximity labeling in vivo. *Neurosci. Res.* **173**, 14–21 (2021).

25. Cijsouw, T. *et al.* Mapping the Proteome of the Synaptic Cleft through Proximity Labeling Reveals New Cleft Proteins. *Proteomes* **6**, 48 (2018).

26. Zhu, F. *et al.* Synapse molecular complexity and the plasticity behaviour problem. *Brain and Neuroscience Advances* **99**, 239821281881068 (2018).

27. Nafstad, P. H. J. An electron microscope study on the termination of the perforant path fibres in the hippocampus and the fascia dentata. *Zeitschrift Für Zellforschung Und Mikroskopische Anatomie* **76**, 532–542 (1967).

28. Montero-Crespo, M. *et al.* Three-dimensional synaptic organization of the human hippocampal CA1 field. *eLife* **9**, e57013 (2020).

29. Amaral, D. G. & Witter, M. P. The three-dimensional organization of the hippocampal formation: A review of anatomical data. *Neuroscience* **31**, 571–591 (1989).

30. Moradi, K. & Ascoli, G. A. A comprehensive knowledge base of synaptic electrophysiology in the rodent hippocampal formation. *Hippocampus* **30**, 314–331 (2020).

31. Reig-Viader, R. & Bayés, A. Quantitative In-Depth Profiling of the Postsynaptic Density Proteome to Understand the Molecular Mechanisms Governing Synaptic Physiology and Pathology. in vol. 127 255–280 (Springer New York, 2017).

32. Goeminne, L. J. E., Gevaert, K. & Clement, L. Experimental design and data-analysis in label-free quantitative LC/MS proteomics: A tutorial with MSqRob. *J Proteomics* **171**, 23–36 (2018).

33. Goeminne, L. J. E., Sticker, A., Martens, L., Gevaert, K. & Clement, L. MSqRob Takes the Missing Hurdle: Uniting Intensity- and Count-Based Proteomics. *Anal Chem* **92**, 6278–6287 (2020).

34. Yao, Z. *et al.* A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *Cell* **184**, 3222-3241.e26 (2021).

35. Lein, E. S. *et al.* Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168–176 (2007).

36. Fabregat, A. *et al.* The Reactome Pathway Knowledgebase. *Nucleic Acids Research* **46**, D649–D655 (2018).

37. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research* **44**, D457-62 (2016).

38. Agrawal, A. *et al.* WikiPathways 2024: next generation pathway database. *Nucleic Acids Res.* **52**, D679–D689 (2023).

39. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).

40. Aleksander, S. A. *et al.* The Gene Ontology knowledgebase in 2023. *GENETICS* **224**, iyad031 (2023).

41. Ulgen, E., Ozisik, O. & Sezerman, O. U. pathfindR: An R Package for Comprehensive Identification of Enriched Pathways in Omics Data Through Active Subnetworks. *Frontiers Genetics* **10**, 858 (2019).

42. Chicco, D. & Agapito, G. Nine quick tips for pathway enrichment analysis. *PLoS Comput. Biol.* **18**, e1010348 (2022).

43. Mubeen, S., Kodamullil, A. T., Hofmann-Apitius, M. & Domingo-Fernández, D. On the influence of several factors on pathway enrichment analysis. *Brief. Bioinform.* **23**, bbac143 (2022).

44. Hanley, J. G. The Regulation of AMPA Receptor Endocytosis by Dynamic Protein-Protein Interactions. *Front Cell Neurosci* **12**, 362 (2018).

45. Lee, S.-J. *et al.* Presynaptic Neuronal Pentraxin Receptor Organizes Excitatory and Inhibitory Synapses. *J. Neurosci.* **37**, 1062–1080 (2016).

46. Yang, Y. & Liu, J.-J. Structural LTP: Signal transduction, actin cytoskeleton reorganization, and membrane remodeling of dendritic spines. *Curr. Opin. Neurobiol.* **74**, 102534 (2022).

47. Narvaes, R. F. & Furini, C. R. G. Role of Wnt signaling in synaptic plasticity and memory. *Neurobiol. Learn. Mem.* **187**, 107558 (2022).

48. McLeod, F. *et al.* Wnt Signaling Mediates LTP-Dependent Spine Plasticity and AMPAR Localization through Frizzled-7 Receptors. *Cell Rep.* **23**, 1060–1071 (2018).

49. Chen, Z. *et al.* Autism-Risk Gene necab2 Regulates Psychomotor and Social Behavior as a Neuronal Modulator of mGluR1 Signaling. *Front. Mol. Neurosci.* **15**, 901682 (2022).

50. Nicodemo, A. A. *et al.* Pyk2 uncouples metabotropic glutamate receptor G protein signaling but facilitates ERK1/2 activation. *Mol. Brain* **3**, 4 (2010).

51. Klaassen, R. V. *et al.* Shisa6 traps AMPA receptors at postsynaptic sites and prevents their desensitization during synaptic activity. *Nature Communications* **7**, 10682 (2016).

52. Ruhl, D. A. *et al.* Synaptotagmin 17 controls neurite outgrowth and synaptic physiology via distinct cellular pathways. *Nat. Commun.* **10**, 3532 (2019).

53. Jurado, S. *et al.* LTP requires a unique postsynaptic SNARE fusion machinery. *Neuron* **77**, 542–558 (2013).

54. Yuan, A. *et al.* Neurofilament subunits are integral components of synapses and modulate neurotransmission and behavior in vivo. *Mol Psychiatr* **20**, 986–994 (2015).

55. Yuan, A. *et al.* Neurofilament light interaction with GluN1 modulates neurotransmission and schizophrenia-associated behaviors. *Transl. Psychiatry* **8**, 167 (2018).

56. Kanai, Y., Dohmae, N. & Hirokawa, N. Kinesin Transports RNA Isolation and Characterization of an RNA-Transporting Granule. *Neuron* **43**, 513–525 (2004).

57. Monday, H. R., Kharod, S. C., Yoon, Y. J., Singer, R. H. & Castillo, P. E. Presynaptic FMRP and local protein synthesis support structural and functional plasticity of glutamatergic axon terminals. *Neuron* **110**, 2588-2606.e6 (2022).

58. Park, H. *et al.* Splice-dependent trans-synaptic PTPδ–IL1RAPL1 interaction regulates synapse formation and non-REM sleep. *EMBO J.* **39**, e104150 (2020).

59. Engelhardt, J. von *et al.* CKAMP44: a brain-specific protein attenuating short-term synaptic plasticity in the dentate gyrus. *Science* **327**, 1518–1522 (2010).

60. Coleman, S. K., Cai, C., Mottershead, D. G., Haapalahti, J.-P. & Keinänen, K. Surface Expression of GluR-D AMPA Receptor Is Dependent on an Interaction between Its C-Terminal Domain and a 4.1 Protein. *J. Neurosci.* **23**, 798–806 (2003).

61. Shen, L., Liang, F., Walensky, L. D. & Huganir, R. L. Regulation of AMPA Receptor GluR1 Subunit Surface Expression by a 4.1N-Linked Actin Cytoskeletal Association. *J. Neurosci.* **20**, 7932–7940 (2000).

62. Lin, D.-T. *et al.* Regulation of AMPA receptor extrasynaptic insertion by 4.1N, phosphorylation and palmitoylation. *Nat. Neurosci.* **12**, 879–887 (2009).

63. Adam, A. R. *et al.* Transcriptional diversity in specific synaptic gene sets discriminates cortical neuronal identity. *Biol Direct* **18**, 22 (2023).

64. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *JMLR* 2825–2830 (2011).

65. Bayés, A. *et al.* Evolution of complexity in the zebrafish synapse proteome. *Nature Communications* **8**, 14613 (2017).

66. Bayés, A. *et al.* Human post-mortem synapse proteome integrity screening for proteomic studies of postsynaptic complexes. *Mol Brain* **7**, 88 (2014).

67. Sanchez-Aguilera, A. *et al.* An update to Hippocampome.org by integrating single-cell phenotypes with circuit function in vivo. *PLoS Biol.* **19**, e3001213 (2021).

68. Wheeler, D. W. *et al.* Hippocampome.org 2.0 is a knowledge base enabling data-driven spiking neural network simulations of rodent hippocampal circuits. *eLife* **12**, (2024).

69. Fidzinski, P., Shor, O. & Behr, J. Target-cell-specific bidirectional synaptic plasticity at hippocampal output synapses. *Eur. J. Neurosci.* **27**, 1111–8 (2008).

70. Alkadhi, K. A. NMDA receptor-independent LTP in mammalian nervous system. *Prog. Neurobiol.* **200**, 101986 (2021).

71. Sindreu, C., Palmiter, R. D. & Storm, D. R. Zinc transporter ZnT-3 regulates presynaptic Erk1/2 signaling and hippocampus-dependent memory. *Proc. Natl. Acad. Sci.* **108**, 3366–3370 (2011).

72. Runge, K., Cardoso, C. & Chevigny, A. de. Dendritic Spine Plasticity: Function and Mechanisms. *Front. Synaptic Neurosci.* **12**, 36 (2020).

73. Favuzzi, E. *et al.* Activity-Dependent Gating of Parvalbumin Interneuron Function by the Perineuronal Net Protein Brevican. *Neuron* **95**, 639-655.e10 (2017).

74. Saroja, S. R. *et al.* Hippocampal proteoglycans brevican and versican are linked to spatial memory of Sprague–Dawley rats in the morris water maze. *J Neurochem* **130**, 797–804 (2014).

75. Zhou, X.-H. *et al.* Neurocan Is Dispensable for Brain Development. *Mol. Cell. Biol.* **21**, 5970–5978 (2001).

76. Frischknecht, R. *et al.* Brain extracellular matrix affects AMPA receptor lateral mobility and short-term synaptic plasticity. *Nat. Neurosci.* **12**, 897–904 (2009).

77. Keller, A., Nesvizhskii, A. I., Kolker, E. & Aebersold, R. Empirical Statistical Model To Estimate the Accuracy of Peptide Identifications Made by MS/MS and Database Search. *Anal. Chem.* **74**, 5383–5392 (2002).

78. Nesvizhskii, A. I., Keller, A., Kolker, E. & Aebersold, R. A Statistical Model for Identifying Proteins by Tandem Mass Spectrometry. *Anal. Chem.* **75**, 4646–4658 (2003).

79. Lein, E. S. *et al.* Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168–176 (2007).

80. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc* **4**, 1184–1191 (2009).

81. Hao, Y. *et al.* Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* 1–12 (2023) doi:10.1038/s41587-023-01767-y.

82. Becht, E. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* **37**, 38–44 (2019).

Methods

Animal handling

All animal research was done with C56BL/6J mice (Jackson Laboratories, Research Resource Identifier, RRID:MGI:5656552) and in accordance with national and European legislation (Decret 214/1997 and RD 53/2013). Research procedures were approved by the Ethics Committee on Animal Research from the Institut de Recerca de l´Hospital de la Santa Creu i Sant Pau (IR-HSCP) and the Departament de Territori i Sostenibilitat from the Generalitat de Catalunya (approval reference num. 9,655). Maintenance and experimental procedures were conducted at the Animal Facility of the IR-HSCP. Mice were housed at a 12h light/dark cycle, with fresh water and food ad libitum. We used animals of both sexes and 9-14 weeks of age. 12 animals were used for laser-capture microdissection proteomics experiments, 2 to isolate postsynaptic density fractions using sucrose gradients and 12 for manual hippocampal dissection and preparation of triton insoluble membranes.

Mouse brain dissection

Mice were culled by cervical dislocation, the head was dissected, and brain removed from skull and meninges. All brain dissection manipulations were done in the presence of chilled 1x phosphate-buffered saline (PBS, 0.144 M NaCl, 2.683 mM KCl, 10.144 mM Na2HPO4, 0.735 mM KH2PO4, [P5368-10PAK from Sigma]). Cerebellum and olfactory bulb were removed prior to any other manipulation. For laser-capture microdissection the forebrain was wrapped in aluminium foil, snap frozen in liquid nitrogen and stored at -80C. For isolation of postsynaptic density (PSD) fractions by ultracentrifugation hippocampi were dissected using iris scissors (PMD120; Thermo Scientific), tissue forceps 1:2 (PMD023445; Thermo Scientific) and scalpel blades in chilled glass petri dishes. Entire hippocampi were frozen at -80C before processing. For manual dissection of CA1, CA3 and DG regions readily dissected hippocampi were first cut coronally in 500 μm slices in the presence of chilled 1x PBS using a tissue slicer (Kerr Scientific Instruments). 8-12 slices where obtained from each hippocampus. Slices were immediately transferred into a glass petri-dish with chilled 1x PBS using a small paint brush. Next CA1, CA3 and DG regions were manually separated from each other using 18G needles (BD) under a microscope Carl Zeiss Meditec model S100 / OPMI 1-FC (see Supplementary Video for a demonstration of manual dissection of hippocampal regions). Dissected regions were placed in individual tubes containing chilled homogenization buffer with phosphatase and protease inhibitors (0,32M Sucrose; 10mM HEPES pH 7,4; 2mM EDTA; 5mM sodium o-vanadate; 30mM NaF; 2μg/ml aprotinin; 2μg/ml leupeptin and 1:2000 PMSF (v/v)) with a pasteur pipette and frozen dry at -80C.

Laser-capture microdissection of neuropil from hippocampal CA3-CA1, DG-CA3 and EC-DG regions

Frozen forebrains were used to obtain 10 μm thick coronal sections in a Leica CM1950 cryostat. Only sections that contained the dorsal hippocampus (Fig. S1a) were processed by laser-capture

microdissection. Sections were placed in membraneSlide 1.0 PEN microscope slides (Zeiss, 415190-9041-000) and stored at -20C. The neuropil of CA1, CA3 and dorsal DG were microdissected using a Leica LMD 6000 laser microdissection microscope. Between 90 and 110 mm^2 were microdissected for each hippocampal region and biological replica. Three biological replicas where generated for each area. All microdissected tissue for each replica was collected in the same 1.5ml tube.

Biochemical isolation of synaptic fractions from laser-capture microdissected tissue

Laser-capture microdissected tissue was collected in 1.5 ml tubes and mixed with PBS containing 1% Triton X-100, 2μg/ml leupeptin and 1/2500 PMSF. The sample was then sonicated in an ultrasonic bath (Branson 1510) for 2 min, incubated in agitation (300rpm) in a ThermoMixer C (Eppendorf) for 30 min at 35C and sonicated again as previously. Afterwards, sample was centrifuged for 10 min at 21.000xg at 4C in a Eppendorf refrigerated centrifuge (5417R). The pellet was resuspended in PBS with 1% SDS. The resuspended pellet and supernatant were mixed with 10x SDS sample buffer for analysis by proteomics or immunoblot. Tissue extraction was also performed with a RIPA buffer containing PBS, 0.1% SDS, 0.5% sodium deoxycholate and 1% Triton X-100.

Biochemical isolation of synaptic enriched fractions from manually dissected hippocampal regions.

Manually dissected hippocampal subregions (CA1, CA3, dDG; see Supplemenatry Video) from 3 animals where accumulated for each biological replica. A total of four biological replicas were prepared for each region. CA1 samples were homogenized in 450μl of homogenizing buffer (HB), CA3 and DG in 300μl. Homogenizing buffer composition: 0,32M Sucrose; 10mM HEPES pH 7,4; 2mM EDTA; 5mM sodium o-vanadate; 30mM NaF; 2μg/ml aprotinin; 2μg/ml leupeptin and 1:2000 PMSF (v/v). Homogenization was performed in 1ml borosilicate tissue homogenizers (357538, Wheaton), using 20-30 strokes. The homogenate was centrifugated in 1.5ml tubes at 800xg and 4C for 10 min in a Eppendorf refrigerated centrifuge (5417R). The pellet, containing the nuclear fraction and cell debris, was re-homogenized once in the same buffer and centrifuged in the same conditions. Supernatants from both centrifugations were pooled and spun down at 10.000xg for 15 min at 4C in the same centrifuge. The resulting pellet was resuspended in Triton buffer (TB: 50mM HEPES pH7.4; 2mM EDTA; 5mM EGTA; 1mM sodium o-vanadate; 30mM NaF; 1% Triton X-100; 2μg/ml aprotinin; 2μg/ml leupeptin and 1:2000 PMSF (v/v)). TB volume used was ½ HB. This mixture was left in ice for 15 minutes and centrifuged at 21.000xg for 30 min at 4C in the same centrifuge. The resulting pellet was resuspended with 30μl of 50mM Tris pH 7.1; 1% SDS and incubated with this buffer for 15 min at room temperature. A final centrifugation was done at 21.000xg for 15 min at room temperature. The resulting supernatant corresponds with the postsynaptic density enriched fraction.

Biochemical isolation of postsynaptic density fractions from whole hippocampus

Isolation of postsynaptic density fractions using standard procedures, which involve the separation of synaptosomes on the bases of their sedimentation rate in sucrose density gradients, was performed as previously described[6,9,65]. Briefly, the hippocampi from two mice were homogenized in 1ml borosilicate tissue homogenizers (357538, Wheaton) adding 9ml of homogenizing buffer for each 1g of tissue weight. Homogenization was done with 20-30 strokes. Homogenizing buffer composed of: 0,32M Sucrose; 10mM HEPES pH 7,4; 2mM EDTA; 5mM sodium o-vanadate; 30mM NaF; 2μg/ml aprotinin; 2μg/ml leupeptin and 1:2000 PMSF (v/v). This sample was first centrifuged at 1400xg and 4C for 10 minutes in an Eppendorf refrigerated centrifuge (5417R). The pellet of this centrifugation was re-homogenized twice following the same procedure. The three supernatants generated were pooled and centrifuged at 700xg for 10 minutes, the pellet was discarded. Next, the sample was centrifuged at 21.000xg for 30 minutes at 4C in the same centrifuge. The resulting pellet was resuspended with Tris 50mM pH7.4 and 0,32M sucrose. A sucrose gradient was prepared with 1 ml of (top to bottom): sample; 50 mM Tris pH 7.4, 0.85 M sucrose; 50 mM Tris pH 7.4, 1 M sucrose; 50 mM Tris pH 7.4, 1.2 M sucrose. This gradient was centrifuged in a SW60Ti rotor (Beckman Coulter) at 82.500xg for 2 hours. The 1.0-1.2 interphase was collected, diluted with 2 equal volumes of 50mM Tris pH 7.4, and centrifuged at 21.000xg for 30 minutes at 4C. The subsequent pellet was resuspended in 50mM Tris pH 7.4, 1% Triton X-100 and maintained in ice for 10 min. This sample was centrifuged at 21,000xg during 30 min at 4C, the resulting pellet corresponds with the fraction enriched with postsynaptic densities.

Protein electrophoresis and Immunoblot

Sample preparation for protein electrophoresis and immunoblot was accomplished by mixing it with 10x SDS loading sample buffer, composition: 500mM Tris pH7.4; 20% SDS; 50% glycerol and 10% b-mercaptoethanol. Prior to its analysis samples were boiled at 95C for 5 min.

SDS-PAGE gels were runed in a vertical MiniProtean system kit (Bio-rad) with 1× running buffer (25 mM TRIS pH 8.4; 0.187 M glycine and 0.1% SDS). Protein standards used were All blue Precision Plus (Bio-Rad). For LC-MS/MS analysis protein gels were stained over night at room temperature with Coomassie solution (B8522-1EA; Sigma-Aldrich) and washed with 2.5% acetic acid and 20% methanol and subsequent washes of 20% methanol, until protein bands were clearly visible. For immunoblot TGX Stain-Free™ gels (161-0181 & 161-0185, SF gels; Bio-Rad) were used and activated as recommended by the manufacturer. Gel images were acquired with ChemiDoc XRS+ (Bio-Rad) and quantified with Image Studio Lite ver. 3.1 (LI-COR Biosciences).

Protein transference was done using a MiniProtean kit (Bio-Rad), and 1× chilled transference buffer (20% methanol; 39 mM Glycine; 48 mM TRIS; 0.04% SDS). Proteins were transferred onto methanol pre-activated polyvinylidene fluoride (PVDF) membranes (IPFL00010, Immobilon-P; Merck-Millipore). Membranes transferred from TGX Stain-Free™ gels were imaged and quantified for posterior normalization with a ChemiDoc XRS+ (Bio-Rad) using the Image Lab

software (Bio-Rad). After transference, PVDF membranes were blocked with 5ml Odissey blocking solution (927-50000; LI-COR) diluted with 1× tris-buffered saline (TBS, 50 mM Tris pH7.4; NaCl 150mM and 0.1% sodium azide). Next, membranes were incubated with primary antibodies in Tween-TBS (T-TBS: 0,1% Tween 20 - TBS) ON at 4C or 1 hour at room temperature. Primary antibodies used: PSD95 (#3450; Cell Signaling, [RRID:AB_2292883]); Synaptophysin (Ab8049; Abcam [SY38], [RRID:AB_2198854]); GluA2 (MAB397; Millipore [RRID:AB_2113875]; Shisa6 (NBP2-85726; Novus Biologicals); mGluR2 (# 191 103; Synaptic Systems [RRID:AB_2232859]; Prkar2a (ab32514; Abcam [RRID:AB_777289]); Ptprd (NBP2-94767; Novus Biologicals). Antibody dilution was 1:1000 except for mGluR2,Ptprd, Prkar2a (1:500) and Shisa6 (1:250). Membranes were washed four times with 1× T-TBS for 5 min before incubation for 1 hr at room temperature protected from light with 5 ml of the following secondary antibodies prepared in T-TBS at a dilution of 1:7.500: anti-rabbit (926-68073, IRDye 680CW, [AB_10954442]), anti-mouse (926-32212, IRDye 800CW [RRID:AB_621847] or 925-68072, IRDye 680RD, [RRID:AB_2814912]) and anti-goat (926-32214, IRDye 800CW, [RRID:AB_621846]). Images were acquired with an Odissey Scanner (LI-COR Biosciences) and protein bands were analyzed with Image Studio Lite ver. 3.1 (LI-COR Biosciences). Protein abundance in postsynaptic density enriched fractions was normalized by the abundance of PSD95, a marker of postsynaptic densities, in order to correct for purity differences between samples.

Sample processing for mass spectrometry analysis

Synaptic fractions obtained from laser-captured microdissected tissue or PSD fractions generated with standard procedures were analysed by conventional protein gel electrophoresis in 6% polyacrylamide gels. For LCM samples gels were runed to half their length and stained with Coomassie as described above. After distaining LCM samples were was cut into 5 bands of the same size (Suppl Fig. 2h). PSD samples were separated into 13 electrophoretic bands (Suppl Fig. 2i). Next, gel bands were cut into 1x1 mm cubes with a scalpel blade in an ethanol cleaned glass plate and under a laminar flow hood. Gel cubes were transfer to 1.5ml tubes for proteomic analysis (0030 123 328; Eppendorf). 50 mM bicarbonate ammonic (BA) in 50% ethanol was added to each tube and incubated for 20 min at room temperature. This solution was replaced with absolute ethanol and incubated 15 more min. For protein reduction gel cubes were mixed with freshly prepared 10mM DTT (dithiothreitol; Merck) in 50mM BA and incubated 1 h at 56C. For protein alkylation, DTT was removed and freshly prepared 55mM IAA (iodacetamide; Merck) in 50mM BA added, incubation was performed in the dark for 30 minutes at room temperature. IAA was removed, 25mM BA added to gel cubes and incubated in the dark for 15 min. For in-gel protein digestion reduced and alkylated samples were mixed with 25 mM BA-50% acetonitrile (ACN) and incubated 15 min twice. Gel cubes were dehydrated with 100% ACN for 10 min. Next, trypsin (Promega) containing solution was prepared and incubated with gel cubes ON at 30C. Tryptic peptides were extracted from gel cubes by first adding 100% ACN and incubating 15 min at 37C. Later, 0.2% trifluoroacetic acid (TFA) was added and incubated for 30 min. Supernatants

were transferred to 0.5 ml tubes (#0030 123 301; Eppendorf) previously washed with ACN to prevent peptide binding to the walls. Liquid-phase was evaporated using a SpeedVac (Thermo-Fisher Scientific). Dried peptides were resuspended in 5% ACN and 0.1% formic acid and bath sonicated for 2 min. Samples were then centrifuged at maximum speed to remove possible gel remainings. Samples were stored at -20C.

Mass spectrometry analysis of tryptic peptides

Tryptic peptides were analysed by LC-MS/MS using an EASY-nLC system (Proxeon Biosystems, Thermo Fisher Scientific) connected to a Velos-Orbitrap mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). Instrument control was performed using Xcalibur software package, version 2.1.0 (Thermo Fisher Scientific, Bremen, Germany). First, peptide mixtures were fractionated by on-line nanoflow liquid chromatography with a two-linear-column system. Digests were loaded onto a trapping guard column (EASY-column, 2 cm long, ID 100 µm, packed with Reprosil C18, 5 µm particle size from Proxeon, Thermo Fisher Scientific) at a maximum pressure of 160 Bar. Then, samples were separated on the analytical column (EASY-column, 10 cm long, ID 75 µm, packed with Reprosil, 3 µm particle size from Proxeon, Thermo Fisher Scientific). Elution was achieved by using a mobile phase from 0.1% formic acid and 100% acetonitrile with 0.1% formic acid and applying a linear gradient from 5 to 35% of buffer B for 120 minutes at a flow rate of 300 nL/min. Ions were generated applying a voltage of 1.9 kV to a stainless-steel nano-bore emitter (Proxeon, Thermo Fisher Scientific), connected to the end of the analytical column. The LTQ Orbitrap Velos mass spectrometer was operated in data-dependent mode. A scan cycle was initiated with a full-scan MS spectrum (from mass to charge [m/z] 300 to 1600) acquired in the Orbitrap with a resolution of 30,000. The 20 most abundant ions were selected for collision-induced dissociation fragmentation in the linear ion trap when their intensity exceeded a minimum threshold of 1000 counts, excluding singly charged ions. Accumulation of ions for both MS and MS/MS scans was performed in the linear ion trap, and the AGC target values were set to 1×10^6 ions for survey MS and 5000 ions for MS/MS experiments. The maximum ion accumulation time was 500 and 200 ms in the MS and MS/MS modes, respectively. The normalized collision energy was set to 35%, and one microscan was acquired per spectrum. Ions subjected to MS/MS with a relative mass window of 10 ppm were excluded from further sequencing for 20 s. For all precursor masses a window of 20 ppm and isolation width of 2 Da was defined. Orbitrap measurements were performed enabling the lock mass option (m/z 445.120024) for survey scans to improve mass accuracy.

LC-MS/MS data was analysed and normalized using Progenesis software (Nonlinear Dynamics, Newcastle, UK). This software allows review of the chromatogram alignments, filtering the data, review peak picking, normalize the data and identify peptides among other features. Specifically, sample ions were automatically aligned to compensate for drifts in retention time between runs. Yet, they were also reviewed and edited manually. The peak picking limits were automatic, the main ion charge selected was set at 4 and the retention time limits were adjusted according the

chromatograms in each sample. Peptide ions were filtered by removing those with a charge of 1 or >4, *m/z* from 300 to 1,600 and the specific retention determined for each case was also set. A normalization step was conducted as it was required to allow comparisons across different sample runs. This normalization was done by assuming that a significant number of peptide ions are unaffected by experimental conditions and the factor by which the sample as a whole varies was used to normalize back to its reference sample in each band from all genotypes analysed.

Database search of mass spectrometry data

All MS/MS samples were analysed using Mascot (Matrix Science, London, UK; version"2.5.1). Mascot was searched with a fragment ion mass tolerance of 0,80 Da and a parent ion tolerance of 10,0 PPM. Charge state deconvolution and deisotoping were not performed. MS/MS spectra were searched with a precursor mass tolerance of 10 ppm, fragment tolerance of 0.5-0.8 Da, trypsin specificity with a maximum of 2 missed cleavages, cysteine carbamidomethylation set as fixed modification (up to 57) and methionine oxidation as variable modification (up to 16). The quantification method applied to quantify protein abundances was a label-free based approach.

Criteria for protein identification by mass spectrometry data

Scaffold (version Scaffold_4.8.5, Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications obtained from Mascot. Peptide identifications were accepted if they could be established at greater than 95,0% probability by the Peptide Prophet algorithm[77] with Scaffold delta-mass correction. Protein identifications were accepted if they could be established at greater than 99,0% probability and contained at least 2 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm[78]. Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony.

Peptide and protein quantification

Peptide abundances were calculated and normalized using Progenesis, which integrates the area under the curve (AUC) of MS1 peaks for peptide quantification. Normalized peptide abundances were exported from Progenesis and peptides from proteins not identified by Scaffold were discarded. Next unique peptides were identified as those defined as non-conflicting by Progenesis or identified as unique by NextProt tool (Expasy) or the Peptide Search tool from Uniprot. Abundances from species of the same unique peptide identified with different retention times were added together. Abundances from modified peptides were added separately. Finally, peptide abundances were normalized based on the average abundance of all peptides from the 14 main postsynaptic density (PSD) scaffolds (Dlg1, Dlg2, Dlg3, Dlg4, Dlgap1, Dlgap2, Dlgap3, Dlgap4, Shank1, Shank2, Shank3, Homer1, Homer2 and Homer3), thus correcting for synaptic enrichment differences between purifications. Peptide abundances were then analysed with MSqROB to obtain protein abundance data and to identify proteins differentially expressed between groups[32,33]. MSqROB was used with the following settings: abundance data was log2

transformed, no normalization was applied, each peptide had to be identified in at least two experiments and only proteins identified with at least 2 peptides were considered for quantification. Furthermore, genotype was used as the fixed effect, while run, sequence and peptide modification were defined as random effects.

Allen Brain Atlas RNA ISH data analysis

Four different scientists manually inspected RNA *in situ* hybridization (ISH) data from adult mouse brain from the Allen Brain Atlas[79]. Each researcher reviewed the 283 proteins overexpressed in CA3-CA1, DG-CA3 and EC-DG synapses. RNA ISH data from the entorhinal cortex was also reviewed for proteins with highest expression in dDG. For a protein to be classified as with highest expression in one or more regions there had to be agreement on 3 out of the 4 researchers. Proteins were classified into those with concordant protein and ISH expression and non-concordant ones. Proteomic data was considered concordant with ISH data when the RNA expression level of a synaptic protein found with highest expression in one of the three hippocampal regions investigated had highest ISH levels in the somas of one or both brain regions contributing to that synapse. For instance, a protein found with highest expression in CA3-CA1 synapses had concordant ISH data if CA3 and/or CA1 somas presented highest expression level of that gene for 3 out of the 4 researchers.

Pathway enrichment analysis

Pathway enrichment analysis was performed using the pathfindR R package[41]. pathfindR takes into consideration protein-protein interaction (PPI) data for pathway enrichment analysis, which is performed using one-sided hypergeometric tests. For our analysis PPI data was retrieved from BioGRID build 4.3.196 (https://thebiogrid.org/) and STRING version 11 (https://string-db.org/), both restricted to *Mus musculus* species. Only STRING interactions with a confidence score above 0.9 were taken into consideration. Redundant interactions between both databases were removed, resulting in a final interaction database with 339.776 interactions. Gene name conversions needed for merging data from different databases and converting them to updated gene symbols were done with biomaRt R package[80]. Pathways investigated with pathfindR were taken from MSigDB collections, (https://www.gsea-msigdb.org) and were restricted to *Mus musculus*. MSigDB contains several collections of gene sets, we used the C2 set: curated gene sets and the C5 set: ontology gene sets. On C2 collection, only REACTOME, WikiPathways and KEGG pathways were used for analysis, which resulted in 2405 gene sets. For the C5 collection all the GO gene sets were selected: Biological process (BP), Cellular Component (CC) and Molecular Function (MF), resulting in 10185 gene sets.

Briefly, pathfindR first builds a Protein Interacting Network (PIN) from all differentially expressed (DE) molecules (genes/proteins) investigated using the PPI data provided. Next, subnetworks are built from the PIN with a minimum length of 10 DE molecules using the Greedy algorithm with a maximum depth of 1, hence only considering the addition of direct neighbours from DE molecules.

Subnetworks with 50% of gene overlap are discarded, maintaining those with a higher score, based on the adjusted p-value of DE molecules. Finally, pathway enrichment analyses is done for each subnetwork, using all the molecules of the PIN as the background set. Pathways that include less than 3 DE molecules are discarded. As the greedy algorithm is a stochastic method, the whole process is repeated 50 times, starting from the subnetwork construction. For a pathway to be considered it had to appear at least in 13 of the 50 (>25%) iterations. Finally, in order to reduce complexity, enriched pathways are grouped using hierarchical clustering, based on their similarity on the DE molecules they include. One 'Representative' term for each cluster was chosen based on the lowest p-value from the hypergeometric test. Heatmaps to represent gene/protein abundance data were generated with the scrattch.hicat R package from the Allen brain atlas (https://github.com/AllenInstitute/scrattch.hicat). Protein and RNA abundance data was normalized by a Log2(x+1) transformation and converted to z-scores.

Source data files relevant to these analysis: Source_Data_6, 7 and 8.

Analysis of single cell RNA-sequencing data from the Allen Brain Cell Atlas

Single cell RNA-seq. data from mouse glutamatergic neurons of the hippocampal formation was retrieved from the Allen Brain Cell Atlas Database (Whole Cortex & Hippocampus - 10X Genomics (2020) with 10X-SMART-SEQ taxonomy[34]). More precisely, we collected RNA-seq. data from the following sub-classes of glutamatergic neurons: DG, CA2-IG-FC, CA3, CA1-ProS, SUB-ProS, CT SUB and NP SUB, all belonging to the hippocampal formation which also includes subiculum neurons[34]. Of note, in this manuscript we refer to ABA Sub-classes as Classes, for simplicity.

Statistical analysis of RNA abundance data was performed using the Seurat R package[81], which is designed to work with single cell gene expression data. To identify DE genes we performed the Wilcoxon Rank Sum test, which is the default test in the Seurat package. p-values were corrected for multiple testing using the Benjamini-Hochberg procedure. As we are interested in identifying abundance differences among genes expressed at synapses, we only worked with RNA abundance data from the genes corresponding to our reference list of synaptic proteins (Suppl. Table 1).

To identify DE genes in a given group (i.e. class or type) we compared gene expression in that group against that of all other groups together. The identification of DE among neuronal types was done within classes. Statistics were done with an equal number of neurons for each group. To identify DE genes between classes we used 100 neurons per class, and to identify DE genes between neuronal types we used 25 neurons per type. In order to sample a representative number of neurons per group so that all DE genes per group would be identified we had to iterate this process. We empirically found that 150 iterations was enough to saturate the number of DE genes in each group. Importantly, for a gene be considered as DE in a given group it had to be identified as significantly DE in at least 90% of these 150 iterations. Furthermore, DE genes not only had

to present and adjusted p-value below 0.05, but their expression fold change value (in log2 scale) had to be above 0.6 for overexpressed genes or below -0.6 for downregulated genes.

Gene expression Dendograms were generated with the median value of log2(x+1) transformed gene expression abundance data and using the scrattch.hicat R package from the Allen brain atlas (https://github.com/AllenInstitute/scrattch.hicat).

Source data files relevant to these analysis: Source_Data_1 to 5.

Uniform Manifold Approximation and Projection (U-MAPS)

To generate neuronal classes and types gene expression U-MAPS we used the umap-learn package (https://pypi.org)[82]. The hyperparameters used to generate the maps were: Random state: 24, Number of neighbours: 15 and Minimum Distance 0.1. All other parameters were left as by default. Only the first two dimensions were used to generate the u-maps.

Gene classification using machine learning

We used the random forest classification method to identify genes with the highest weight in the organization of neurons in classes and types. Gene expression data from the Allen Brain atlas was analysed with the 'Random Forest Classifier' function within the scikit-learn (https://scikit-learn.org/0.16/about.html) Python package[64]. The hyperparameters used for the Random Forest Classifier were: Random state: 24, Max. Depth: 12 and Number of estimators: 200. Values for all other parameters were kept as by default. The test set used included 20% of neurons in each group and the train set the remaining 80%. The 'confusion matrix' function from scikit-learn was used to generate confusion matrices.

Source data file relevant to these analysis: Source_Data_9.

Source Data:

Source_Data_1_Iteration_Classes.R: R script to iterate the statistical analysis performed with Seurat to identify genes differentially expressed between neuronal classes.

Source_Data_2_Iteration_Types.R: R script to iterate the statistical analysis performed with Seurat to identify genes differentially expressed between neuronal types.

Source_Data_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.

Source_Data_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.

Source_Data_5_Split_Types.R: R script to obtained data from a subset of neuronal types from the entire transcriptomic database provided by the ABCA.

Source_Data_6_pathfindR_Proteomics.Rmd: R script to perform the pathfinder analysis and to generate the heatmaps from the proteomics data.

Source_Data_7_pathfindR_Classes.R: R script to perform the pathfinder analysis and to generate the heatmaps from transcriptomics data of neuronal classes (ABCA).

Source_Data_8_pathfindR_Types.R: R script to perform the pathfinder analysis and to generate the heatmaps from transcriptomics data of neuronal types (ABCA).

Source_Data_9_ Random_Forest.ipynb: Python code to perform the Random Forest analysis on transcriptomic data from the ABCA.

Acknowledgments

RRV, DdCB and AB financial support was provided by: PID2021-124411OB-I00 and RTI2018-097037-B-I00 (MINECO/MCI/AEI/FEDER, EU), Award AC17/00005 by ISCIII through AES2017 and within the NEURON framework, Ramón y Cajal Fellowship (RYC-2011-08391p), IEDI-2017-00822 and AGAUR (2017 SGR 1776 and 2021 SGR 01005). DdCB thanks AGAUR/Generalitat de Catalunya/FEDER, EU for 'Ajuts per a la contractació de personal investigador novel (FI)' Ref.2020FI_B00130. All authors thank the CERCA Programme/Generalitat de Catalunya for institutional support.

Author Contributions

RRV, DdCB, ABP, DAA, DRV performed experiments. CS and AB designed and supervised all experiments and secured funding. AB and CS wrote the manuscript. All authors reviewed and approved the manuscript.

Competing interests:

Authors declare no competing interests.