RESEARCH

Transcriptional diversity in specific synaptic gene sets discriminates cortical neuronal identity

Amparo Roig Adam[1,2†] , José A. Martínez-López[1†], Sophie J. F. van der Spek[3], The SYNGO consortium, Patrick F. Sullivan[4,5], August B. Smit[3] , Matthijs Verhage[2,6*] and Jens Hjerling-Leffler[1*]

Abstract

Synapse diversity has been described from different perspectives, ranging from the specific neurotransmitters released, to their diverse biophysical properties and proteome profiles. However, synapse diversity at the transcriptional level has not been systematically identified across all synapse populations in the brain. To quantify and identify specific synaptic features of neuronal cell types we combined the SynGO (Synaptic Gene Ontology) database with single-cell RNA sequencing data of the mouse neocortex. We show that cell types can be discriminated by synaptic genes alone with the same power as all genes. The cell type discriminatory power is not equally distributed across synaptic genes as we could identify functional categories and synaptic compartments with greater cell type specific expression. Synaptic genes, and specific SynGO categories, belonged to three different types of gene modules: gradient expression over all cell types, gradient expression in selected cell types and cell class- or type-specific profiles. This data provides a deeper understanding of synapse diversity in the neocortex and identifies potential markers to selectively identify synapses from specific neuronal populations.

Introduction

Synapses are the information processing units of the brain and function in a use-dependent manner [1, 2]. Synapses are diverse with regards to subcellular targets and physiological properties [3] and are central in information processing and storage theories [4]. Moreover, brain regions involved in higher cognitive functions, such as the hippocampus and neocortex, contain greater synapse diversity [5]. Synapse diversity also overlaps with the connectivity patterns (connectome)

between brain areas associated to different functions [5]. Thus, understanding synapse diversity is crucial for gaining insights into the mechanisms for information processing in the brain.

Neurotransmitters and biophysical properties have traditionally been used for classification of synapses [2, 4]. New technologies, including diverse '*omics*', are now uncovering additional layers of complexity and diversity on the molecular signatures of synapses [6]. Proteome differences between synapse types correlate with functional diversity (strength, kinetics, or synaptic plasticity) [4]. These different molecular profiles include, for example, scaffold proteins PSD95 and SAP102 [5, 7], and AMPA-type glutamate receptors (AMPARs) [4] for postsynaptic terminals of excitatory cells; and Gephyrin (*GPHN*) and Collybistin (*ARHGEF9*) scaffold proteins, and the $GABA_A$ receptors ($GABA_A$R) for inhibitory postsynaptic sites [8]. On the

[†]Amparo Roig Adam and José A. Martínez-López contributed equally to this work

*Correspondence:
Matthijs Verhage
matthijs@cncr.vu.nl
Jens Hjerling-Leffler
jens.hjerling-leffler@ki.se
Full list of author information is available at the end of the article

Roig Adam *et al. Biology Direct*   (2023) 18:22

Page 2 of 10

presynaptic side, synaptotagmins 1 and 2, involved in calcium-dependent vesicle exocytosis, are differentially expressed between synapse types [9]. While most previous work is centered on inhibitory and excitatory synapses, recent studies have pointed out the expression of synaptic genes as correlated with neuronal diversity in subpopulations of transcriptomic cell types [10, 11].

Here, we aim to systematically identify how synaptic gene expression specifies the diversity of neuronal cell types in mouse neocortex using single cell transcriptomics data. Combining the expert-curated, evidence-based SynGO synaptic ontology [12] and single cell expression data we could observe that expression of synaptic genes presents a striking diversity. Remarkably, specific biological function and synaptic component gene categories contained significantly high diversity and discrete modules of synaptic genes exhibited different modes of variability revealing that synapse diversity is organized at different levels.

Methods
Datasets
The single cell RNA-sequencing dataset published by Tasic et al. [13] together with the information available in the SynGO [12] database were retrieved for the study of the expression of synaptic genes in the neocortex. The scRNA-seq dataset used was generated using the smart-seq RNA-sequencing technique. The tissue used was mouse primary visual cortex and anterior lateral motor cortex and up to 133 transcriptomic cell types and 16 cell classes were identified. The described cell classes include glutamatergic neurons, labeled according to their preferential layer of residence (for example Layer 4 neurons are labelled as L4) and their projection pattern (intratelencephalic, IT; pyramidal tract, PT; near-projecting, NP; and corticothalamic, CT) and GABAergic neurons labelled with their predominant expressed gene including: *Sst*, *Pvalb*, *Vip*, *Lamp5*, *Sncg* and *Serpinf1* [13]. The first release of the expert-curated synaptic gene ontology, SynGO1.0 was used [12]. SynGO1.0 contains 1112 unique human genes annotated to 2918 terms hierarchically organized and divided into 'Cellular Components' and 'Biological Processes' related to the synapse. These annotated synaptic genes encode evidenced proteins that localize to synaptic compartments and contribute to synaptic functions.

Pre-processing and visualization
The scRNA-seq data were filtered to keep only the cells belonging to the classes 'GABAergic' and 'Glutamatergic' ($n = 22{,}439$ cells) and expression data of the synaptic genes included in SynGO (1049 genes). Three subsets of the data were used for the downstream analysis separately by filtering the genes in the original dataset according to the following gene sets: all synaptic genes present in the SynGO database, SynGO presynaptic genes and SynGO postsynaptic genes.

Each of the filtered datasets was pre-processed with the standard protocol used by the Seurat [14] R package as follows: log-normalization and scaling (scaling factor = 10,000) of the raw count data, identification of highly variable genes, PCA dimensionality reduction, selection of significant principal components (PCs) by the Jackstraw procedure, and tSNE (t-distributed Stochastic Neighbor Embedding) dimensionality reduction/visualization (perplexity = 50). The significant PCs determined for each data subset were: 60 PCs for the full dataset, 42 PCs for the dataset with all synaptic genes and 21 PCs for both datasets with the pre- and postsynaptic genes. No quality filtering was performed on the cells since the dataset used was already passing the quality criteria in Tasic et al. [13]. The tSNE embedding of the data were color coded with the cluster identities determined by Tasic et al. [13].

Synaptic function and localisation (SynGO annotations) underlying diversity
MetaNeighbor [15] was used to measure the power of each of the SynGO annotations to discriminate between different cell types. For each gene set (or SynGO annotated term), AUROC (Area Under the Receiver Operator Curve) scores were calculated for each of the 16 described cell types. To do so, random samples of the dataset were taken to train (2/3) the algorithm and test (1/3) the gene sets. Only those gene sets with at least 2 genes were used in this analysis. The result is an AUROC score for each gene set that can be interpreted as the performance of the gene set for the task of identifying each cell type, with 0.5 being equivalent to a random guess.

To calculate the statistical significance of the AUROC scores, the performance of random gene sets in MetaNeighbor was compared to that obtained with the SynGO annotations by generating random gene sets of the same size. For each gene set size in SynGO 10,000 gene sets were generated by sampling from all the genes expressed in the original dataset, as well as all genes found in SynGO. For each of these randomly generated gene sets the 'fast_version' of MetaNeighbor was used. Firstly, the AUROC scores were used to compare the average performance of random gene sets and random synaptic gene sets of each size. Secondly, the random synaptic gene sets were used to calculate the statistical significance of the SynGO annotation scorings by calculating an empirical *p* value. As indicated in Eq. 1,

Roig Adam *et al. Biology Direct*   (2023) 18:22

Page 3 of 10

this is done by calculating the fraction of scores from the randomized gene sets that are higher than the scores of the SynGO annotations. To calculate this as the overall score across all cell types, the score used for the *p* value calculation was the sum of the AUROC scores in the 16 cell types. Likewise, we calculated the empirical *p* value of SynGO annotations performing significantly worse than random gene sets (fraction of AUROC scores from randomized synaptic gene sets lower than the scores of SynGO annotations).

$$pval = \frac{\#\left\{ \sum random\ gene\ sets\ AUROC\ \geq \sum SynGO\ AUROC \right\}}{N};$$
(1)

where N = total number of random permutations (10,000)

To calculate the specificity per gene in Additional file 1: Table S1 we ranked the synaptic genes according to the cell type specificity of their expression by calculating the proportion of expression of each gene in each cell type similarly to Skene et al. [16]. To do so, we first normalized the gene expression in each cell type by aggregating the counts for each gene across cells belonging to the same cell type, scaling to 1 million counts and then dividing by the total counts in all cells in that cell type. Then the specificity score was calculated by dividing the normalized expression of each gene in every cell type by the total expression if the gene in every cell type. The list of synaptic genes was then ranked by the maximum score of each gene, indicating that top ranked genes have a higher cell type specificity.

Quantification of cell type diversity encoded by synaptic genes

To measure and compare the cell type diversity observed with different gene sets MetaNeighbor analysis was performed as described in the previous section. The quantified gene sets included the most high variable genes among: all genes in the dataset, non-synaptic genes (defined as all genes excluding the genes in SynGO), all synaptic genes, presynaptic genes, postsynaptic genes and mitochondrial genes (all genes included in the dataset and annotated in MitoCarta [17]). AUROC scores were calculated for each gene set and cell type, as well as for every cell class. Wilcoxon rank test (followed by false discovery rate [FDR] correction) was used to determine statistically different performance of each pair of gene sets.

Synapse gene correlation network analysis

Weighted gene correlation network analysis (WCGNA) was used to investigate modules of synaptic genes in the transcriptional network of the dataset. In brief, the standard pipeline from the WGCNA [18] R package was used to perform hierarchical clustering on the distance between every gene pair, calculated as 1-TOM (topological overlap matrix). To generate the TOM matrix, the co-expression similarity matrix is raised to a soft thresholding power (adjacency) that approximates a scale-free topology while keeping the mean connectivity of the network ($\beta = 4$) [19]. Finally, the clustered genes are grouped into modules of highly interconnected genes using a dynamic branch-cutting algorithm. We used the dynamic tree cutting function (maximum height 0.9, 0.95, 0.98) and the modules were selected from a consensus of the result.

The gene modules were classified using K-means clustering on the eigenvector that explains the variance of gene expression in each cell type (80.3% variability explained). To do so, the average gene expression of each module in each cell was used to calculate the variance of expression for each gene module in each cluster. Next, the cell type identity information was removed, and the variance matrix ordered. The eigenvector explaining the maximum variability in the data (PC1) was used to cluster the modules in groups of similar variances of gene expression.

The individual gene modules were characterized using two approaches: mapping the average expression of the module to the synaptic types and annotating the function or cellular compartment they are related to by gene ontology enrichment. The former was done by mapping the average expression of all genes in each module, normalized to the average expression of random genes, to the transcriptomic cell types and visualizing it in the tSNE generated using only synaptic genes. To map the function and cellular component most related to each gene module, hypergeometric gene set enrichment was used. The background used for this analysis (universe) was comprised by all the genes annotated in SynGO that were present in the Tasic et al. [13] dataset. The significance scores (*p* value) from the hypergeometric tests were adjusted for multiple hypothesis testing using the Bonferroni correction method. Lastly, visualization of the test results for every gene module was produced with the sunburst custom color-coding tool of SynGO ontologies [12].

Results

Synapse genes contain cell identity information

To evaluate transcriptional diversity of synaptic genes in neuronal cell types, we filtered the expression data of cell types identified by Tasic et al. [13] using the genes in SynGO. We analyzed four gene sets, including all genes in the original dataset (Fig. 1A), all synaptic genes in SynGO (Fig. 1B), presynaptic genes in SynGO (Fig. 1C) and postsynaptic genes in SynGO (Fig. 1D). We then

Roig Adam *et al. Biology Direct*   (2023) 18:22

Page 4 of 10

compared the cell class and cell type diversity across the four different subsets. Here, we refer to the 133 transcriptomic neuronal types described in Tasic et al. [13] as cell types (different colors in Fig. 1) and 16 merged groups of these cell types as cell classes. Distinct classes and cell types could be discerned using SynGO genes only and all genes in the dataset to a similar extent (Fig. 1A, B). Additionally, the observed transcriptomic diversity of presynaptic (Fig. 1C) and postsynaptic (Fig. 1D) genes showed similar levels of cell type specification. Quantification of the class and cell type discriminatory power of synaptic gene expression was calculated as the cell classification performance of each gene set using MetaNeighbor [15]. We included a similar sized gene-set from Mito-Carta [17] as comparison. Synaptic genes had a similar power in discerning classes and cell types in comparison to all genes or after removing SynGO genes (Fig. 1E, F). We observed no difference in the discriminatory power between presynaptic and postsynaptic gene sets (Fig. 1E, F). Similarly, no difference was observed when measuring pre- and post-synaptic genes in excitatory and inhibitory neurons independently (Additional file 3: Fig. S1A). However, there is a considerable overlap between the terms pre- and post-synaptic genes. Comparing only the genes specific for each category revealed a significantly higher score of postsynaptic genes for inhibitory neurons (Additional file 3: Fig. S1B, C). This suggests that postsynaptic diversity is larger among GABAergic cells. These results indicate that the diversity of the synapse transcriptome across cell types is similar to that of the full transcriptome. In Additional file 1: Table S1 we provide individual specificity scores (see methods) for each of the synaptic genes in the analysis.

Specific SynGO annotations underlie synapse diversity

To identify whether genes contributing to synapse diversity belong to specific functional sets or are expressed in specific synaptic compartments, we analyzed the cell type discriminatory power of annotated SynGO terms. To test this, we used MetaNeighbor [15] to score the performance of each SynGO term on the task of discriminating different cell types (AUROC scores) and compared it to random sets (of equivalent size) of genes drawn from SynGO and from all expressed genes in the dataset (Fig. 2, Additional file 4: Fig. S2A). Several SynGO terms in both biological functions (BP, Fig. 2A,

B) and cellular components (CC, Fig. 2C, D) discriminated cell types significantly better than random gene sets. Among the top biological process annotations are elements of the postsynaptic density organization, synaptic signaling, modulation of presynaptic chemical transmission and synaptic vesicle exocytosis. For cellular localisation, both presynaptic and postsynaptic membranes, as well as the presynaptic cytosol and active zone membrane were significant. A few categories conversely performed worse than random, including ribosomal genes and genes involved in metabolism (Additional file 4: Figure S2B, C). Analysis of average expression per category could not explain this result (Additional file 4: Fig. S2D, E). This analysis confirmed that synaptic genes perform better, on average, in cell type identification analysis than gene sets comprised of any gene expressed in the data also when normalising for number of genes. These results show that synapse diversity among different neuronal types accumulates in specific functions and cellular components.

Gene network analysis reveals different levels of synaptic organisation

WGCNA analysis and hierarchical clustering of the gene co-expression network revealed a high level of modularity of synaptic genes (Fig. 3A, Additional file 2: Table S2). Classification of the gene modules according to the eigenvector calculated from the variance of gene expression across cell types (Fig. 3B), showed that synaptic gene modules can be clustered into three types: modules with specific expression in cell types or cell classes (discrete modules), modules showing a gradient of expression in a specific cell class (intermediate gradients) and modules with a similar gradient of expression in all cell types (pure gradients). Notably, these different classes of diversity were similarly found in pre- and postsynaptic gene modules. The results from this analysis were mapped to cell types using the average expression of the gene module in each cell (Fig. 3E, G, I; Additional file 5: Fig. S3). Interestingly, we observed modules with cell type specific expression in *Vip*-cells, sometimes shared with other cell types including *Sncg*-cells (pink; Fig. 3I) and near projecting cells (dark green; Additional file 5: Fig. S3). This suggest that some synaptic specializations can be re-used between GABAergic cell types and across

(See figure on next page.)

Fig. 1 Neuronal diversity is recovered using only synaptic genes. t-SNE embedding visualization of the dataset from Tasic et al. [13] using the most variable genes among **A** all genes in the dataset, **B** only synaptic genes annotated in SynGO, **C** only presynaptic genes or **D** only postsynaptic genes, allow distinction of the annotated cell types to a similar extent. Dotted lines indicate cell classes and colors correspond to cell types described in Tasic et al. [13]. Quantification was performed by calculating the cell class (**E**) and cell type (**F**) discriminatory power of each gene set in the MetaNeighbor pipeline. Wilcoxon rank test was used to determine statistically different performance of each pair of gene sets (*: $p < = 0.05$; **: $p < = 0.01$; ***: $p < = 0.001$; ****: $p < = 0.0001$)
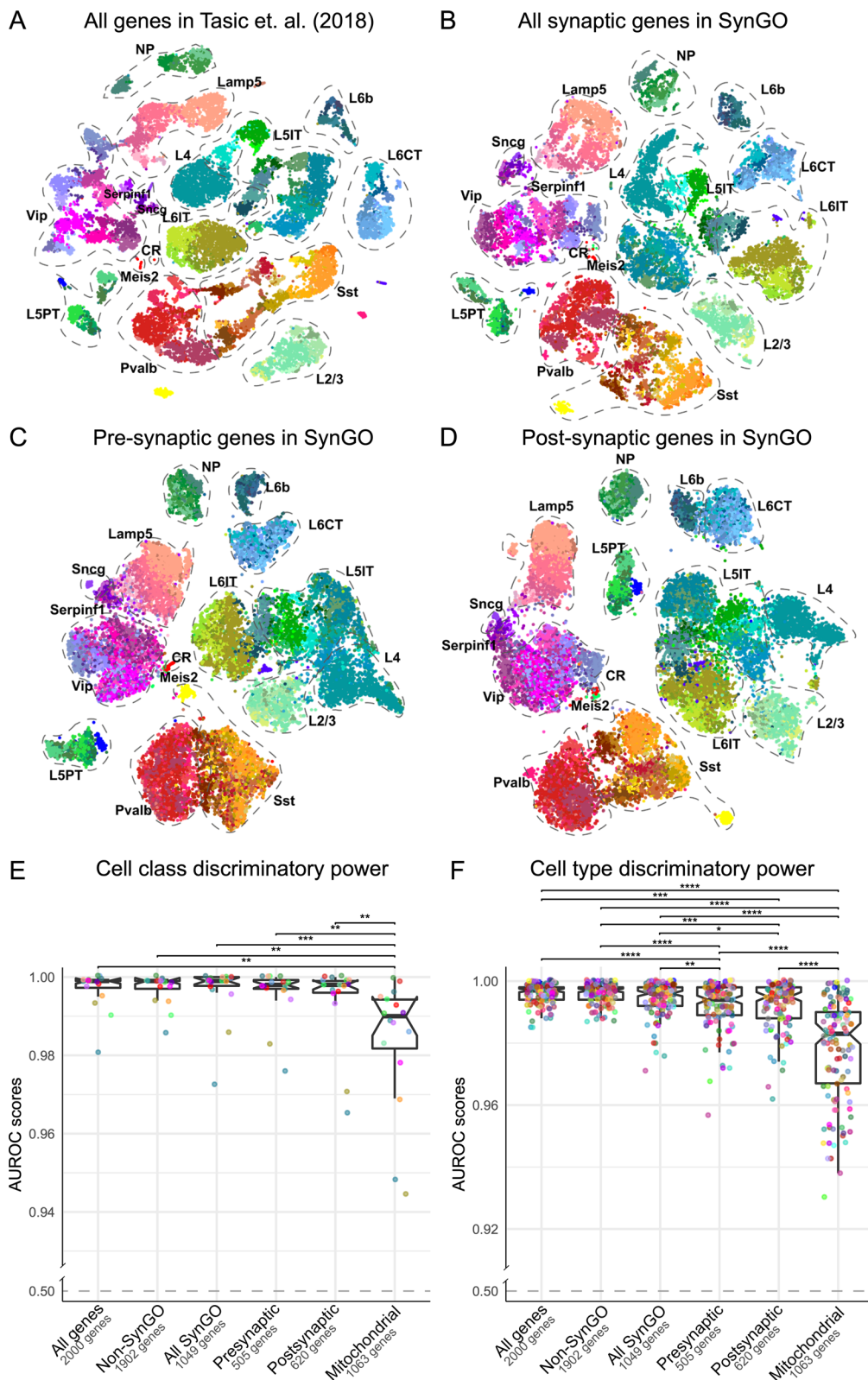
Fig. 1 (See legend on previous page.)

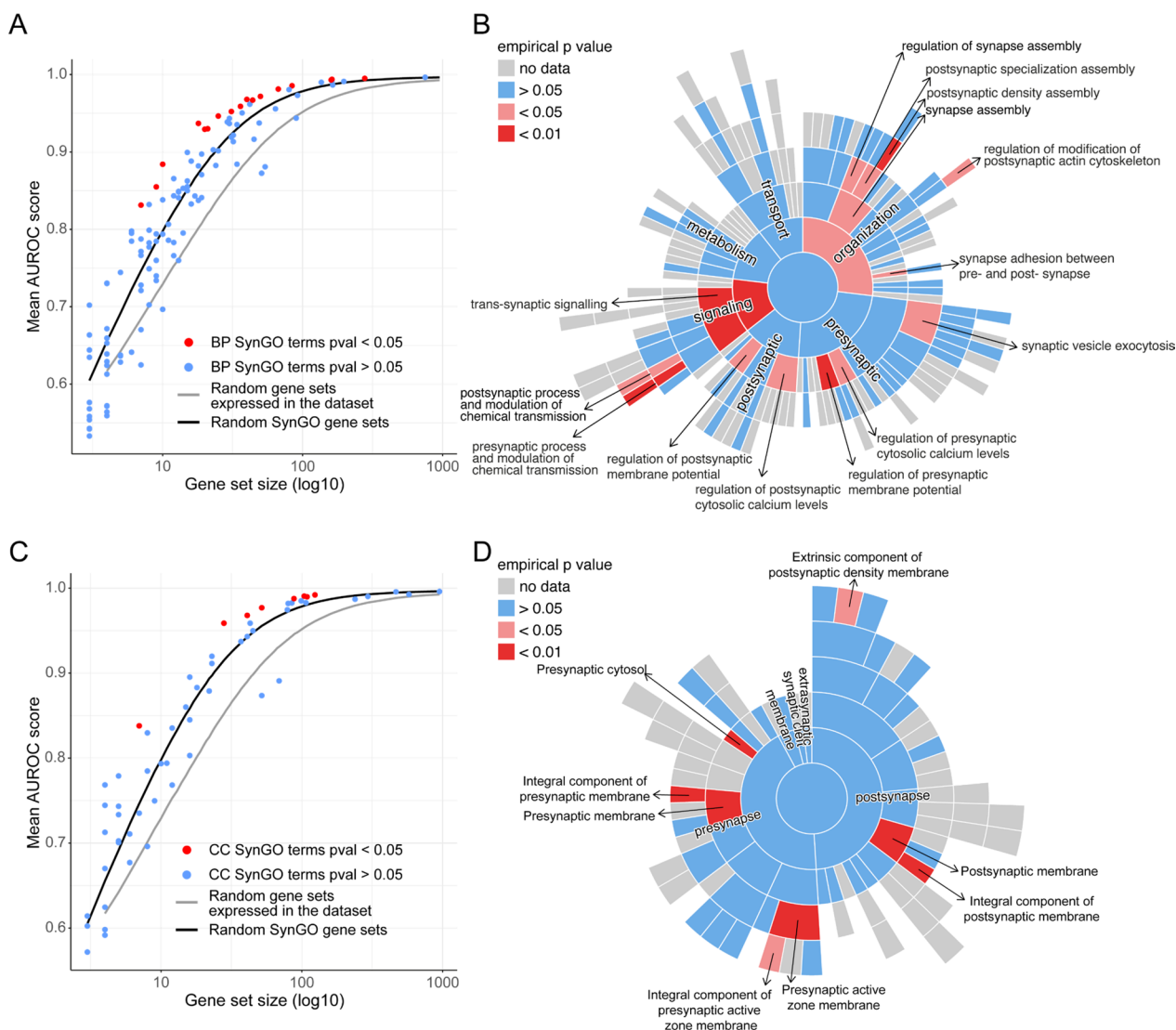Roig Adam *et al. Biology Direct*   (2023) 18:22

Page 6 of 10

Fig. 2 Synapse diversity resides in specific functions and cellular compartments. For all SynGO categories, the mean AUROC score across the 16 cell types is shown for biological functions (**A**) and cellular compartments (**C**) annotated in SynGO. Some SynGO terms (red) perform better than random synaptic gene sets of the same size (black line). Randomly generated synaptic gene sets (black line) discriminate cell types better than random gene sets (grey line) regardless the set size (Wilcoxon rank sum test; W = 3049, p = 0.001). The sunburst plots show the SynGO biological processes (**B**) and cellular compartments (**D**) where most variability lies across all neuronal subclasses. The color code (p value) indicates SynGO terms that perform significantly better than random synaptic gene sets of the same size

GABAergic and excitatory cell types. In addition, gene set enrichment analysis of the obtained gene modules showed the biological processes and cell compartments (SynGO terms) to which each gene module is most related (Fig. 3D, F, H). Interestingly, none of the enriched SynGO terms in the different groups of modules are overlapping between groups. Modules exhibiting gradients of expression included terms related to metabolism, post- and pre-synaptic ribosome, and protein translation (similar to those terms indicated in Additional file 4: Fig. S2B). Interestingly we observed

two gradient modules with opposing expression pattern (Fig. 3C) suggesting that these are specific programs that are anti-regulated, perhaps in response to external signals or each other. This included the genes *CTBP1* and *ARL8* involved in "presynapse to nucleus signaling pathway" and "regulation of anterograde synaptic vesicle transport respectively" (yellow module), opposing the expression pattern of ribosomal and translational machinery genes (turquoise module). These results suggest that there are different types of synaptic organisation, ranging from cell type-specific
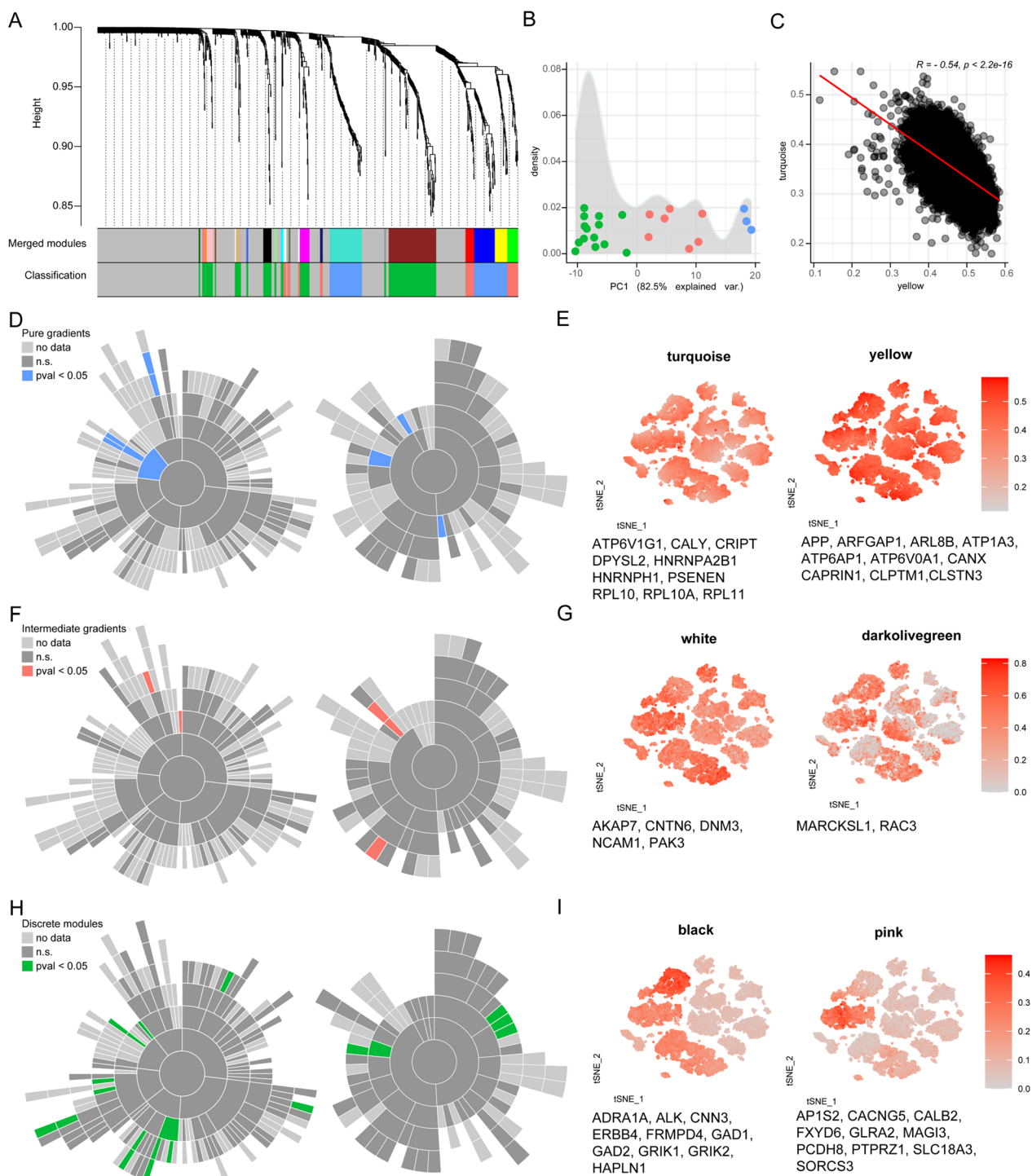
Roig Adam *et al. Biology Direct*   (2023) 18:22

Page 7 of 10

Fig. 3 WGCNA reveals different levels of synaptic organisation: pure gradients within cell types, intermediate gradients and discrete expression in specific cell types and cell classes. **A** WGCNA dendrogram and gene modules selected. **B** Density distribution of the eigenvector (PC1) that explains the variance of gene expression across cell types in the gene modules (80.3% variance explained). Colour coding corresponds to the K-means gene module classification according to their gradients across cell types. **C** Anticorrelation of the average gene expression of the turquoise and yellow gradient modules. **D**, **F**, **H** Sunburst plots showing the biological functions (left) and cellular components (right) for which the gene modules of each type show enrichment. Dark grey indicates non-significant SynGO terms that contain genes in the modules, and coloured SynGO terms indicate enrichment in one of the gene modules. **E**, **G**, **I** tSNE plot of example modules within each group colour coded with the average gene expression of the genes in each module

Roig Adam *et al. Biology Direct*   (2023) 18:22

Page 8 of 10

to pan-neuronal programs, specified by distinct sets of genes at the transcriptome level, which also involve specific cellular functions.

Discussion

In this study, synapse diversity was mapped to previously defined transcriptomic neuronal cell types. We found that synaptic genes contain considerable cell identity information at the transcriptome level. Among synaptic genes, certain groups of genes associated to specific synaptic functions and localisation, annotated as SynGO terms, underlie the observed synapse diversity. Moreover, we identified additional candidate modules of coexpressed genes that contribute to synaptic functional diversity. These gene modules suggest different types of synapse organisation or different hierarchies of synaptic specification.

These results agree with the proposed vast synapse diversity arising from the combination of the different proteins that have been described as part of the synapse proteome [4, 20]. Therefore, transcriptomic synapse diversity exists to a deeper extent of that depicted by the classical classifications of synapse types, possibly integrating the anatomical and physiological features classically described, as proposed for GABAergic interneurons in previous studies [11].

We observed cell type-related diversity in both the pre- and postsynaptic genes. Our findings add additional gene-level resolution to the postsynaptic site diversity previously proposed in the brain based on protein expression of *Dlg4* (PSD95) and *Dlg3* (SAP102) [5]. Additionally, our data highlights the existence of such diversity also in the presynaptic site, showing a similar molecular diversity.

Our results show that synapse diversity, as well as similarity, between different cell types resides in specific synaptic functions and components. We identified cytoskeleton organisation, cell adhesion and synaptic signaling, as important for synapse diversity. As expected, we observed gene modules specific to excitatory/inhibitory synapse classification but also gene modules being specific to neuronal classes and neuronal types. An additional layer of diversity seems to be related to gradient-like expression of gene modules within each cell type, and surprisingly gene modules showing opposing expression which is likely an indication of dynamic synapse regulation as proposed by Zu et al. [5].

Despite the single-neuron synapse diversity depicted here, recent studies have also described synapse diversity within a single neuron [6]. Differential spatial distribution of synapse mRNA and proteins across the dendritic tree or between the cell body and synapses likely represent distinct functions within the same cell. It is our hope that our results broaden the understanding of synapse diversity and generate hypotheses for future single synapse research. Revealing the subcellular localization of these mRNA and proteins can provide insights on the synapse diversity within one neuron and the dynamic processes that occur in response to activity, perhaps through local translation of proteins. As an example, gene modules showing gradient expression profiles within cell types could reflect different cell states of the same cell types, in which single synapse variability could have a role. Our study provides the opportunity to expand the knowledge on the specific synaptic profile of distinct cell types. Further work in this direction could be used to selectively identify populations of synapses derived from specific populations of neuronal cell types, in intact tissue as well as in disease models.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13062-023-00372-y.

Additional file 1. Table S1. Ranked list of synaptic genes according to their maximum cell type specificity.

Additional file 2. Table S2. List of genes in each of the modules identified with WGCNA.

Additional file 3. Figure S1. Comparison of cell type discriminatory power of pre and post synaptic genes within cell classes. Quantification cell type identity encoded in each all genes annotated in SynGO, all pre- and post-synaptic genes; and non-overlapping pre- and post-synaptic genes, as well as the tSNE embedding resulting from only exclusive pre- and post-synaptic genes are shown. Wilcoxon rank test was used to determine statistically different performance of each pair of gene sets. Colour code of each cell type is the same used in Tasic et al.and Fig 1A.

Additional file 4. Figure S2. All MetaNeighbor AUROC scores obtained in the random gene sets used for bootstrap analysis and annotated SynGO categories for each cell class in the dataset.Mean AUROC score across the 16 cell types is shown for biological functionsand cellular compartmentsannotated in SynGO. Some SynGO termsscore significantly worse than the random performance expected for their respective gene set size. The sunburst plots show the SynGO biological processesand cellular compartmentswhere less variability than expected lies across all neuronal subclasses. The colour codeindicates SynGO terms that perform significantly worse than random synaptic gene sets of the same size.Average expression of all genes in each annotated SynGO term.

Additional file 5. Figure S3. tSNE representation of the synaptic cell typescolour coded with the average expression of the genes in each module found with WGCNA.

Roig Adam *et al. Biology Direct*   (2023) 18:22

Page 9 of 10

Tilmann Achsel (Department of Fundamental Neurosciences, University of Lausanne, CH-1006 Lausanne, Switzerland). Maria Andres-Alonso (Leibniz Group 'Dendritic Organelles and Synaptic Function', Center for Molecular Neurobiology, ZMNH, University MC, Hamburg-Eppendorf, Hamburg, 20251, Germany; RG Neuroplasticity, Leibniz Institute for Neurobiology, 39118 Magdeburg, Germany). Claudia Bagni (Dept. of Fundamental Neurosciences, University of Lausanne, CH-1006 Lausanne, Switzerland; Department of Biomedicine and Prevention, University of Rome Tor Vergata, 00133 Rome, Italy). Àlex Bayés (Molecular Physiology of the Synapse Laboratory, Institut d'Investigació Biomèdica Sant Pau (IIB SANT PAU), Sant Quintí 77-79, 08041 Barcelona, Spain; Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain). Thomas Biederer (Department of Neurology, Yale School of Medicine, New Haven, CT 06511, USA). Nils Brose (Department of Molecular Neurobiology, Max Planck Institute of Multidisciplinary Sciences, 37075 Göttingen, Germany). John Jia En Chua (Department of Physiology, Yong Loo Lin School of Medicine, National University of Singapore; LSI Neurobiology Programme, National University of Singapore; Healthy Longevity Translational Research Program, Yong Loo Lin School of Medicine, National University of Singapore; Institute of Molecular and Cell Biology, Agency for Science, Technology and Research (A*STAR), Singapore). Marcelo P. Coba (Zilkha Neurogenetic Institute, Department of Psychiatry and Behavioral Sciences and Department of Physiology and Neuroscience, Keck School of Medicine, University of Southern California, Los Angeles, CA 90333, USA). L. Niels Cornelisse (Functional Genomics section, Department Human Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam University Medical Center, De Boelelaan 1085, 1081HV Amsterdam, The Netherlands). Jaime de Juan-Sanz (Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, Inserm, CNRS, APHP, Hôpital de la Pitié Salpêtrière, Paris, France). Hana L. Goldschmidt (Solomon H. Snyder Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD, 21205, USA; Kavli Neuroscience Discovery Institute, Johns Hopkins University, Baltimore, MD 21205, USA). Eckart D. Gundelfinger (Leibniz Institute for Neurobiology (LIN), Brenneckestraße 6, 39118 Magdeburg, Germany; Center for Behavioral Brain Sciences (CBBS) and Institute of Pharmacology and Toxicology, Medical Faculty, Otto von Guericke University, 39120 Magdeburg, Germany). Richard L. Huganir (Solomon H. Snyder Department of Neuroscience Kavli Neuroscience Discovery Institute The Johns Hopkins University School of Medicine Baltimore, MD 21205; Kavli Neuroscience Discovery Institute, Johns Hopkins University, Baltimore, MD 21205, USA). Cordelia Imig (Department of Neuroscience, University of Copenhagen, Blegdamsvej 3B, 2200 Copenhagen N, Denmark). Reinhard Jahn (Laboratory of Neurobiology, Max-Planck Institute for Biophysical Chemistry, 37077 Göttingen, Germany; University of Goettingen, 37077 Goettingen, Germany). Hwajin Jung (Center for Synaptic Brain Dysfunctions, Institute for Basic Science (IBS), Department of Biological Sciences, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, South Korea). Pascal S. Kaeser (Department of Neurobiology, Harvard Medical School, Boston, MA 02115, USA). Eunjoon Kim (Center for Synaptic Brain Dysfunctions, Institute for Basic Science (IBS), Daejeon 34141, South Korea; Department of Biological Sciences, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, South Korea). Frank Koopmans (Department of Functional Genomics, Center for Neurogenomics and Cognitive Research, Vrije Universiteit Amsterdam, 1081 HV Amsterdam, The Netherlands; Department of Molecular and Cellular Neurobiology, Center for Neurogenomics and Cognitive Research, Vrije Universiteit Amsterdam, 1081 HV Amsterdam, The Netherlands). Michael R. Kreutz (RG Neuroplasticity, Leibniz Institute for Neurobiology, 39118 Magdeburg, Germany; Leibniz Group 'Dendritic Organelles and Synaptic Function', Center for Molecular Neurobiology, ZMNH, University MC, Hamburg-Eppendorf, Hamburg, 20251, Germany). Noa Lipstein (Department of Molecular Physiology and Cell Biology, Leibniz-Forschungsinstitut für Molekulare Pharmakologie, Robert-Rössle-Str. 10, 13125 Berlin, Germany). Harold D. MacGillavry (Cell Biology, Neurobiology and Biophysics, Faculty of Science, Utrecht University Kruytgebouw, room N504, Padualaan 8, 3584 CH Utrecht, The Netherlands). Peter S. McPherson (Department of Neurology and Neurosurgery, Montreal Neurological Institute, McGill University, Montreal, Quebec H3A 2B4, Canada). Vincent O'Connor (Biological Sciences, University of Southampton, Southampton, SO17 1BJ, UK). Rainer Pielot (Institute for Pharmacology and Toxicology, Medical Faculty Otto-von-Guericke University Magdeburg, Leipziger Strasse 44, 39120 Magdeburg, Germany; CBBS Magdeburg, Germany). Timothy A. Ryan (Department of Biochemistry, Weill Cornell Medicine, New York, NY 10065, USA). Carlo Sala (CNR Neuroscience Institute Milan, 20854 Vedano al Lambro (MB), Italy). Morgan Sheng (Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, 75 Ames Street, Cambridge, MA 02142, USA; Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology (MIT), 43 Vassar St, Cambridge, MA 02139, USA). Karl-Heinz Smalla (Institute for Pharmacology and Toxicology, Medical Faculty Otto-von-Guericke University Magdeburg, Leipziger Strasse 44, 39120 Magdeburg, Germany; Leibniz Institute for Neurobiology (LIN), Brenneckestraße 6, 39118 Magdeburg, Germany; CBBS Magdeburg, Germany; Center for Behavioral Brain Sciences (CBBS) and Institute of Pharmacology and Toxicology, Medical Faculty, Otto von Guericke University, 39120 Magdeburg, Germany). A. B. Smit (Department of Molecular and Cellular Neurobiology, Center for Neurogenomics and Cognitive Research, Vrije Universiteit Amsterdam, 1081 HV Amsterdam, The Netherlands). Paul D. Thomas (Division of Bioinformatics, Department of Population and Public Health Sciences, Keck School of Medicine of USC, University of Southern California, Los Angeles, CA 90033, USA). Ruud F. Toonen (Department of Functional Genomics, Center for Neurogenomics and Cognitive Research, Vrije Universiteit Amsterdam, 1081 HV Amsterdam, The Netherlands). Jan R. T. van Weering (Functional Genomics section, Department Human Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam University Medical Center, De Boelelaan 1085, 1081 HV Amsterdam, The Netherlands). Matthijs Verhage (Department of Functional Genomics, Center for Neurogenomics and Cognitive Research, Vrije Universiteit Amsterdam, 1081 HV Amsterdam, The Netherlands; Functional Genomics section, Department Human Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam University Medical Center, De Boelelaan 1085, 1081HV Amsterdam, The Netherlands). Chiara Verpelli (CNR Neuroscience Institute Milan, 20854 Vedano al Lambro, MB, Italy)

Funding

Availability of data and materials
The datasets analysed in this study were previously published and are available at GSE115746 and https://www.syngoportal.org/

Code availability
Custom scripts used in this study can be found at: https://github.com/Hjerling-Leffler-Lab/SynGO_scRNAseq

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Division of Molecular Neurobiology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, 17177 Stockholm, Sweden. [2]Present Address: Department of Functional Genomics, Center for Neurogenomics and Cognitive Research (CNCR), Amsterdam Neuroscience, Vrije Universiteit Amsterdam, 1081 HV Amsterdam, The Netherlands. [3]Department of Molecular and Cellular Neurobiology, Center for Neurogenomics and Cognitive Research (CNCR), Amsterdam Neuroscience, Vrije Universiteit Amsterdam, 1081 HV Amsterdam, The Netherlands. [4]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [5]Department of Genetics, University of North Carolina, Chapel Hill, NC 27599-7264, USA. [6]Functional Genomics Section, Department of Human Genetics, Center for Neurogenomics and Cognitive Research (CNCR), Amsterdam University Medical Center (UMC), De Boelelaan 1085, 1081 HV Amsterdam, The Netherlands. [7]Department of Fundamental Neurosciences, University

of Lausanne, 1006 Lausanne, Switzerland. [8]Leibniz Group 'Dendritic Organelles and Synaptic Function', Center for Molecular Neurobiology, ZMNH, University MC, Hamburg-Eppendorf, 20251 Hamburg, Germany. [9]RG Neuroplasticity, Leibniz Institute for Neurobiology, 39118 Magdeburg, Germany. [10]Department of Biomedicine and Prevention, University of Rome Tor Vergata, 00133 Rome, Italy. [11]Molecular Physiology of the Synapse Laboratory, Institut d'Investigació Biomèdica Sant Pau (IIB SANT PAU), Sant Quintí 77-79, 08041 Barcelona, Spain. [12]Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain. [13]Department of Neurology, Yale School of Medicine, New Haven, CT 06511, USA. [14]Department of Molecular Neurobiology, Max Planck Institute of Multidisciplinary Sciences, 37075 Göttingen, Germany. [15]Department of Physiology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. [16]LSI Neurobiology Programme, National University of Singapore, Singapore, Singapore. [17]Healthy Longevity Translational Research Program, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. [18]Institute of Molecular and Cell Biology, Agency for Science, Technology and Research (A*STAR), Singapore, Singapore. [19]Department of Psychiatry and Behavioral Sciences and Department of Physiology and Neuroscience, Keck School of Medicine, Zilkha Neurogenetic Institute, University of Southern California, Los Angeles, CA 90333, USA. [20]Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, Inserm, CNRS, APHP, Hôpital de La Pitié Salpêtrière, Paris, France. [21]Solomon H. Snyder Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA. [22]Kavli Neuroscience Discovery Institute, Johns Hopkins University, Baltimore, MD 21205, USA. [23]Leibniz Institute for Neurobiology (LIN), Brenneckestraße 6, 39118 Magdeburg, Germany. [24]Center for Behavioral Brain Sciences (CBBS), Institute of Pharmacology and Toxicology, Medical Faculty, Otto Von Guericke University, 39120 Magdeburg, Germany. [25]Solomon H. Snyder Department of Neuroscience Kavli Neuroscience Discovery Institute, The Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA. [26]Department of Neuroscience, University of Copenhagen, Blegdamsvej 3B, 2200 Copenhagen N, Denmark. [27]Laboratory of Neurobiology, Max-Planck Institute for Biophysical Chemistry, 37077 Göttingen, Germany. [28]University of Goettingen, 37077 Goettingen, Germany. [29]Department of Biological Sciences, Center for Synaptic Brain Dysfunctions, Institute for Basic Science (IBS), Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, South Korea. [30]Department of Neurobiology, Harvard Medical School, Boston, MA 02115, USA. [31]Center for Synaptic Brain Dysfunctions, Institute for Basic Science (IBS), Daejeon 34141, South Korea. [32]Department of Biological Sciences, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, South Korea. [33]Department of Functional Genomics, Center for Neurogenomics and Cognitive Research, Vrije Universiteit Amsterdam, 1081 HV Amsterdam, The Netherlands. [34]Department of Molecular and Cellular Neurobiology, Center for Neurogenomics and Cognitive Research, Vrije Universiteit Amsterdam, 1081 HV Amsterdam, The Netherlands. [35]Department of Molecular Physiology and Cell Biology, Leibniz-Forschungsinstitut Für Molekulare Pharmakologie, Robert-Rössle-Str. 10, 13125 Berlin, Germany. [36]Cell Biology, Neurobiology and Biophysics, Faculty of Science, Utrecht University Kruytgebouw, Room N504, Padualaan 8, 3584 CH Utrecht, The Netherlands. [37]Department of Neurology and Neurosurgery, Montreal Neurological Institute, McGill University, Montreal, QC H3A 2B4, Canada. [38]Biological Sciences, University of Southampton, Southampton SO17 1BJ, UK. [39]Institute for Pharmacology and Toxicology, Medical Faculty, Otto-Von-Guericke University Magdeburg, Leipziger Strasse 44, 39120 Magdeburg, Germany. [40]CBBS, Magdeburg, Germany. [41]Department of Biochemistry, Weill Cornell Medicine, New York, NY 10065, USA. [42]CNR Neuroscience Institute Milan, 20854 Vedano al Lambro, MB, Italy. [43]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, 75 Ames Street, Cambridge, MA 02142, USA. [44]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology (MIT), 43 Vassar St, Cambridge, MA 02139, USA. [45]Division of Bioinformatics, Department of Population and Public Health Sciences, Keck School of Medicine of USC, University of Southern California, Los Angeles, CA 90033, USA.

References

1. Abbott LF, Regehr WG. Synaptic computation. Nature. 2004;431:796–803. https://doi.org/10.1038/nature03010.
2. Jackman SL, Regehr WG. The mechanisms and functions of synaptic facilitation. Neuron. 2017;94:447–64. https://doi.org/10.1016/j.neuron.2017.02.047.
3. Kubota Y, Karube F, Nomura M, Kawaguchi Y. The diversity of cortical inhibitory synapses. Front Neural Circuits. 2016;10:27. https://doi.org/10.3389/fncir.2016.00027.
4. O'Rourke NA, Weiler NC, Micheva KD, Smith SJ. Deep molecular diversity of mammalian synapses: why it matters and how to measure it. Nat Rev Neurosci. 2012;13:365–79. https://doi.org/10.1038/nrn3170.
5. Zhu F, Cizeron M, Qiu Z, et al. Architecture of the mouse brain synaptome. Neuron. 2018;99:781-799.e10. https://doi.org/10.1016/J.NEURON.2018.07.007.
6. Grant SGN, Fransén E. The synapse diversity dilemma: molecular heterogeneity confounds studies of synapse function. Front Synaptic Neurosci. 2020;12:45. https://doi.org/10.3389/fnsyn.2020.590403.
7. Broadhead MJ, Bonthron C, Arcinas L, et al. Nanostructural diversity of synapses in the mammalian spinal cord. Sci Rep. 2020;10:8189. https://doi.org/10.1038/s41598-020-64874-9.
8. Crosby KC, Gookin SE, Garcia JD, et al. Nanoscale subsynaptic domains underlie the organization of the inhibitory synapse. Cell Rep. 2019;26:3284-3297.e3. https://doi.org/10.1016/j.celrep.2019.02.070.
9. Südhof TC. The synaptic vesicle cycle. Annu Rev Neurosci. 2004;27:509–47. https://doi.org/10.1146/annurev.neuro.26.041002.131412.
10. Zeisel A, Hochgerner H, Lönnerberg P, et al. Molecular architecture of the mouse nervous system. Cell. 2018;174:999-1014.e22. https://doi.org/10.1016/J.CELL.2018.06.021.
11. Paul A, Crow M, Raudales R, et al. Transcriptional architecture of synaptic communication delineates GABAergic neuron identity. Cell. 2017;171:522-525.e20. https://doi.org/10.1016/j.cell.2017.08.032.
12. Koopmans F, van Nierop P, Andres-Alonso M, et al. SynGO: an evidence-based, expert-curated knowledge base for the synapse. Neuron. 2019;103:217-234.e4. https://doi.org/10.1016/j.neuron.2019.05.002.
13. Tasic B, Yao Z, Graybuck LT, et al. Shared and distinct transcriptomic cell types across neocortical areas. Nature. 2018;563:72–8. https://doi.org/10.1038/s41586-018-0654-5.
14. Butler A, Hoffman P, Smibert P, et al. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol. 2018;36:411–20. https://doi.org/10.1038/nbt.4096.
15. Crow M, Paul A, Ballouz S, et al. Characterizing the replicability of cell types defined by single cell RNA-sequencing data using MetaNeighbor. Nat Commun. 2018;9:884. https://doi.org/10.1038/s41467-018-03282-0.
16. Skene NG, Bryois J, Bakken TE, et al. Genetic identification of brain cell types underlying schizophrenia. Nat Genet. 2018;50:825–33. https://doi.org/10.1038/s41588-018-0129-5.
17. Rath S, Sharma R, Gupta R, et al. MitoCarta3.0: an updated mitochondrial proteome now with sub-organelle localization and pathway annotations. Nucleic Acids Res. 2021;49:D1541–7. https://doi.org/10.1093/nar/gkaa1011.
18. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinform. 2008;9:559. https://doi.org/10.1186/1471-2105-9-559.
19. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol. 2005. https://doi.org/10.2202/1544-6115.1128.
20. Grant SGN. Toward a molecular catalogue of synapses. Brain Res Rev. 2007;55:445–9. https://doi.org/10.1016/J.BRAINRESREV.2007.05.003.

Publisher's Note

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