

Human brain prefrontal cortex proteomics identifies compromised energy metabolism and neuronal function in Schizophrenia

Frank Koopmans¹, Anke A. Dijkstra^{2,3}, Wei-Ping Li^{1,4}, Remco V. Klaassen¹, Yvonne Gouwenberg¹, Shuyang Yao⁵, Lisa Bast⁶, Matthijs Verhage^{7,8}, Robert Karlsson⁵, Andrew J. Dwork⁹, Craig A. Stockmeier¹⁰, Jens Hjerling-Leffler⁶, Patrick F. Sullivan^{5,11}, August B. Smit^{1*}

1. Department of Molecular and Cellular Neurobiology, Center for Neurogenomics and Cognitive Research (CNCR), Vrije Universiteit (VU) Amsterdam, Amsterdam, The Netherlands.
2. Department of Pathology, Amsterdam University Medical Center (UMC), Amsterdam, The Netherlands.
3. Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, Netherlands.
4. Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research (CNCR), Vrije Universiteit (VU) Amsterdam, Amsterdam, The Netherlands.
5. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden.
6. Department of Medical Biochemistry and Biophysics, Division of Molecular Neurobiology, Karolinska Institutet, Stockholm 17177, Sweden.
7. Department of Functional Genomics, Center for Neurogenomics and Cognitive Research (CNCR), Vrije Universiteit (VU) Amsterdam, Amsterdam, The Netherlands.
8. Department of Human Genetics, Amsterdam University Medical Center (UMC), Amsterdam, The Netherlands.
9. Department of Pathology and Cell Biology, Columbia University, New York, NY, USA.
10. Department of Psychiatry and Human Behavior, University of Mississippi Medical Center, Jackson, MS, USA.
11. Department of Genetics, University of North Carolina, Chapel Hill, NC, 27599-7264, USA.

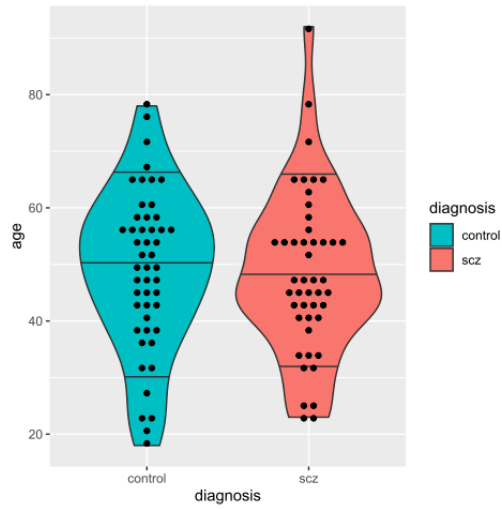


Figure S1. Age distributions of control- and patient-donors are highly similar.

The median age for $N=49$ controls (green) and $N=47$ patients (red) is 50 and 47.5 years old, respectively. Data available in Supplementary Data 1.

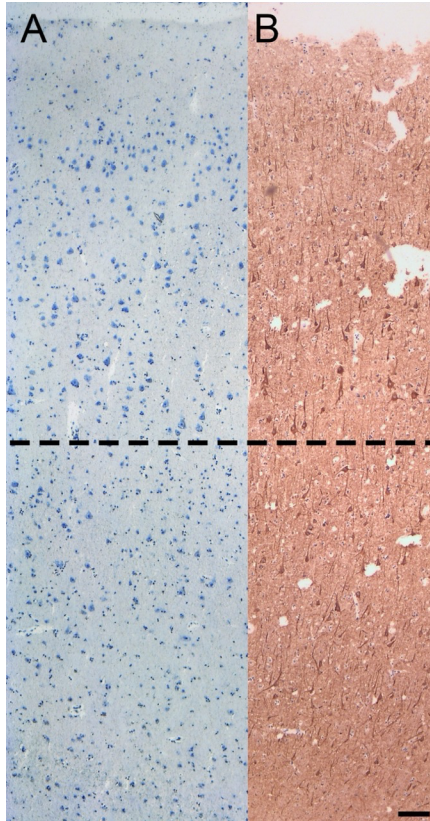


Figure S2. Cortical lamination of sections used for LCM.

Based on the Toluidine blue (A), Layer 4 could be determined for most cases by its granular appearance. *SMI32* (B) staining further confirmed the lower border of Layer 3 where large *SMI32*-positive neurons are present. The lower border of Layer 6 was determined by the Toluidine blue staining. Scale bar A,B = 100 μ m.

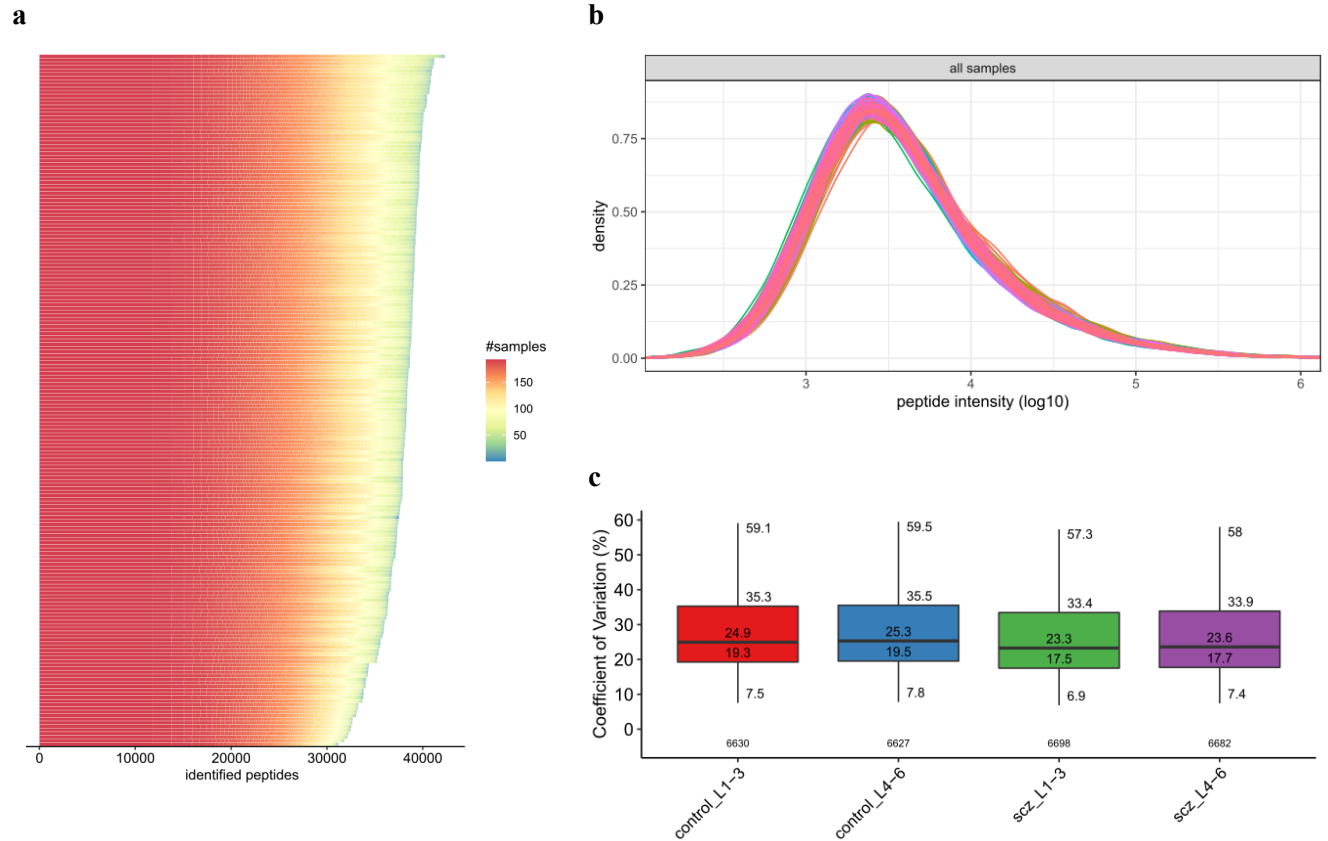


Figure S3. Quality control analyses generated by MS-DAP.

a) Identified peptide counts over all samples, color coded by the number of detections made in other samples (blue indicates peptides that are rarely detected throughout the entire dataset). **b)** Peptide intensity distributions show strong overlap for all samples, indicating that the protein sample quantities loaded into the mass-spectrometer were highly consistent. **c)** Distributions of the protein-level Coefficient of Variation (CoV) across replicates within each experimental condition. Protein counts (N) used for CoV computation per group are shown below each boxplot. The Layer sets, within control and schizophrenia groups, are abbreviated as L1-3 and L4-6.

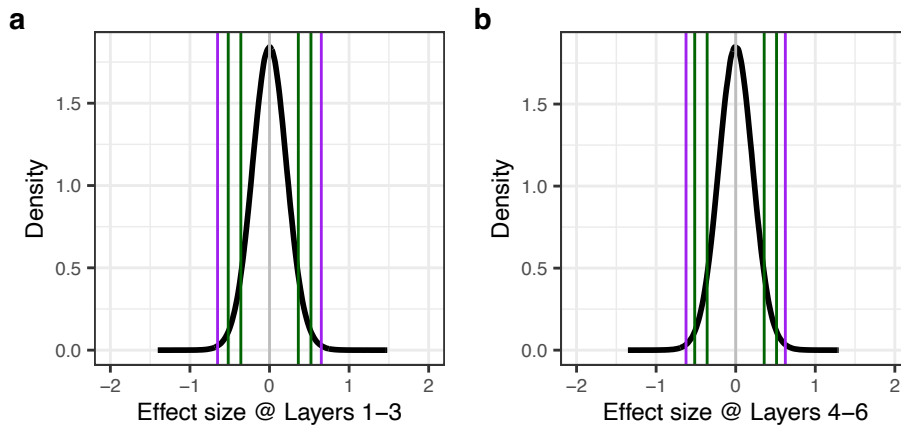


Figure S4. Bootstrapping of control versus schizophrenia regression analyses.

Sample labels (metadata) were randomly permuted and linear regression analyses (Methods) were repeated across 1000 bootstrap iterations to generate distributions of estimated protein effect sizes for the control versus schizophrenia comparison. Distributions are shown for Layers 1-3 (**a**) and Layers 4-6 (**b**). Green vertical lines indicate the 1%, 5%, 95% and 99% quantiles of the bootstrapped distributions. Purple vertical lines indicate the smallest absolute effect sizes (within up- or down-regulation) observed in the actual linear regression model for the subset of proteins significant at FDR adjusted p-value 0.05 (Fig. 1b-c, Supplementary Data 2), i.e. the primary results reported in the manuscript.

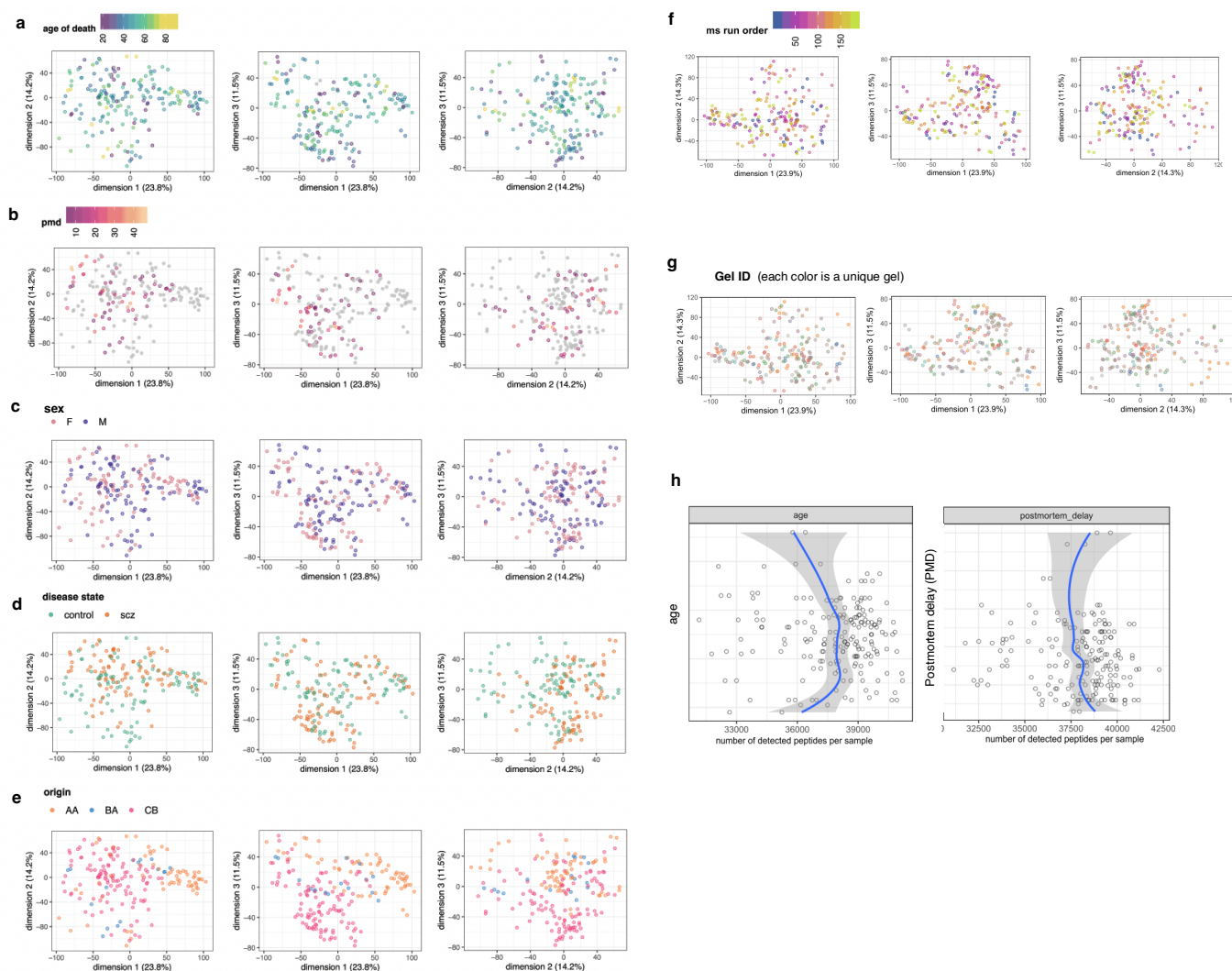


Figure S5. Quality control analyses for potential confounders generated by MS-DAP.

a-g) PCA plots of various covariates, including potential confounders of the analysis. Each dot represents an individual sample. Samples cluster only according to their diagnosis (**d**) and origin (**e**, source where the samples are obtained from; Stockmeier (CB), Netherlands Brain Bank (BA), Dwork (AA)). Postmortem delay (PMD) information was only available for a subset of samples (grey dots). **h)** Detection of the number of peptides is not dependent on age (of death) or postmortem delay (PMD). MS: mass spectrometer.

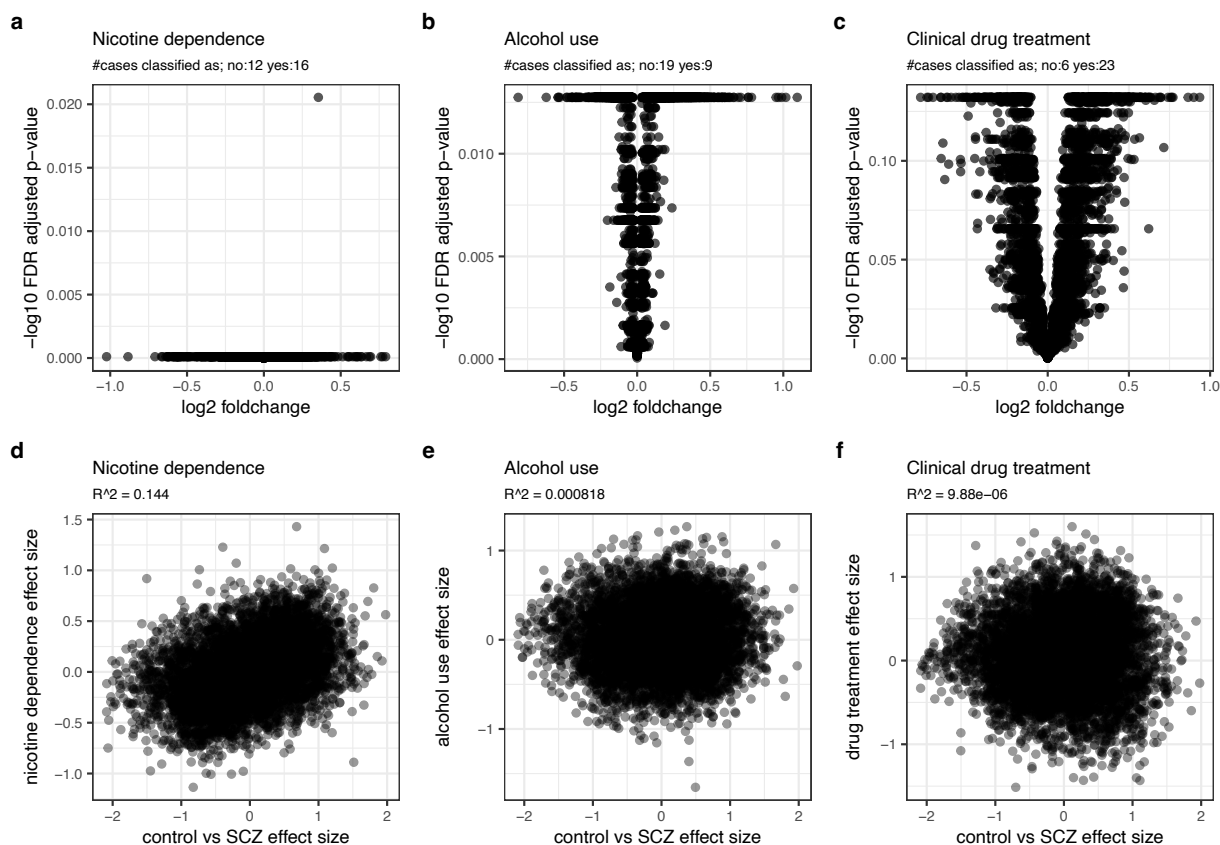


Figure S6. Statistical analysis of potential confounding effects.

a) Differential expression analysis was performed using limma linear regression models, with DEqMS correction for peptide counts per protein, to assess the impact of nicotine dependence (yes or no) on protein abundance for samples within the Stockmeier cohort ($N=56$). Both Layer sets were included in the analyses, accounting for repeated measurements from the same individual using a block design (Methods). Protein log₂ foldchanges are shown on x-axis, FDR adjusted p-values are shown on the y-axis. **b-c)** analogous analyses for alcohol use (yes or no) and drug treatment (yes or no), respectively. Note that after p-value adjustment for multiple testing, p-values are far from significant as evidence from the y-axis of panels **a-c**. Sample metadata are described in Supplementary Data 1. **d-f)** Correlation plots showing effect sizes from the analyses from panels **a-c**, against effect sizes from control versus SCZ analyses (also across both Layer sets), further supporting the lack of confounding by these variables. Source data are provided as a Source Data file.

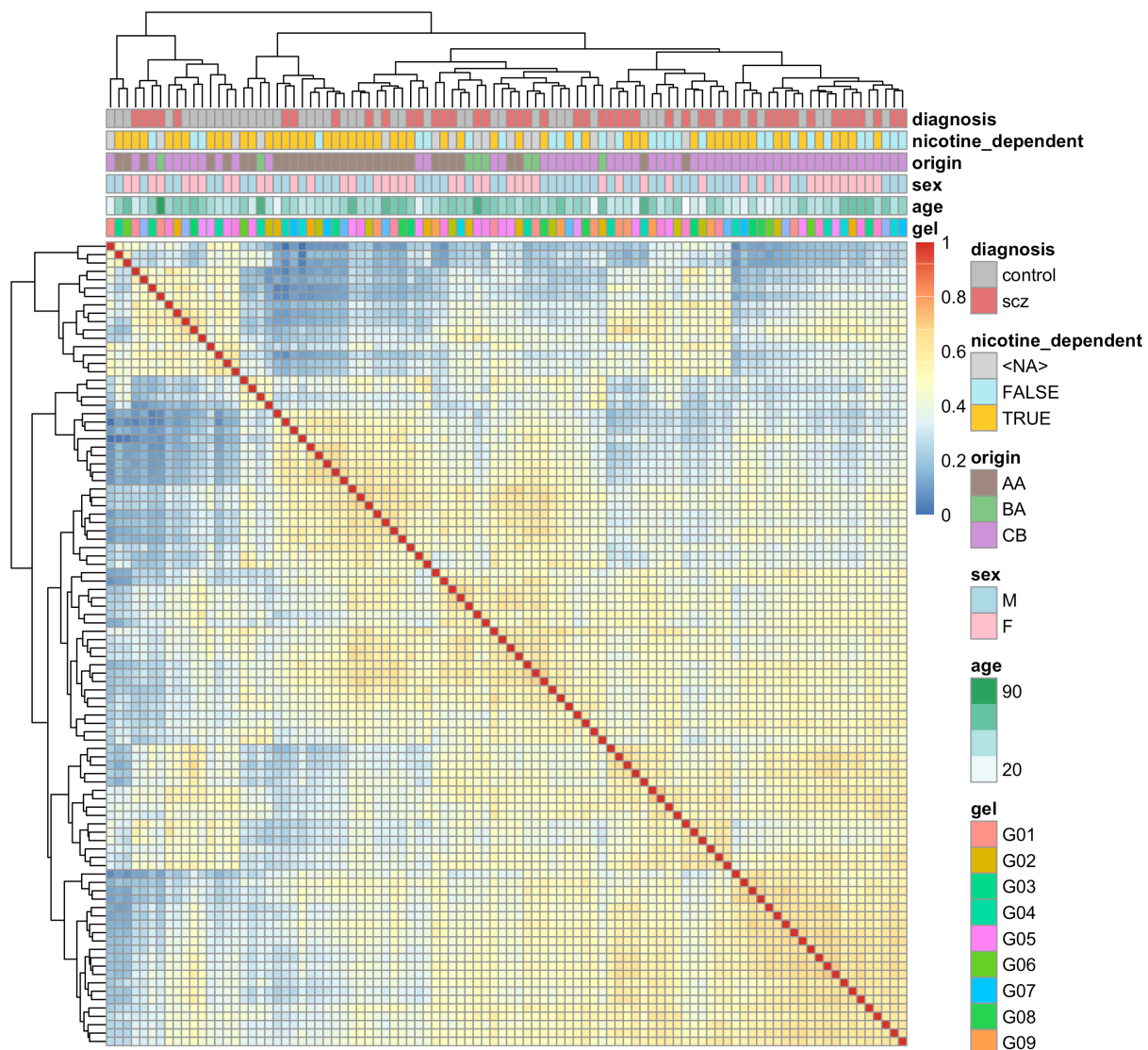


Figure S7. Unsupervised clustering of Layer 1-3 samples across all individuals.

The protein abundance (log2 intensity) matrix for all samples from Layers 1-3 was used to compute the sample similarity (1 - Euclidean distance) matrix. This sample*sample similarity matrix was used to perform unsupervised clustering analysis of individuals using the pheatmap R package with default options. Colors indicate sample similarity scores between pairs of samples. Annotated sample metadata from Supplementary Data 1 is shown as color-coded bars on top; diagnosis (control or schizophrenia/SCZ), nicotine dependence (TRUE, FALSE, Not Available), age, cohort (the source where samples were collected), sex (Male, Female), age, gel (included to check for a potential technical batch in sample preparation). Only 1 strong cluster is observed (top-left), which could not be related to any of the annotated metadata and is a mix of diagnoses; we observe no apparent SCZ patient subclusters. Source data are provided as a Source Data file.

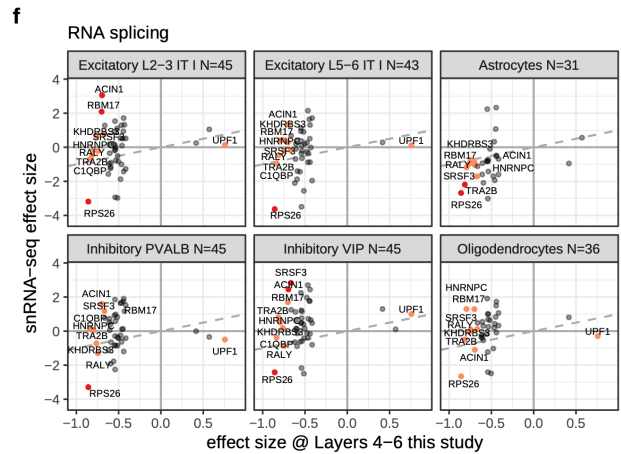
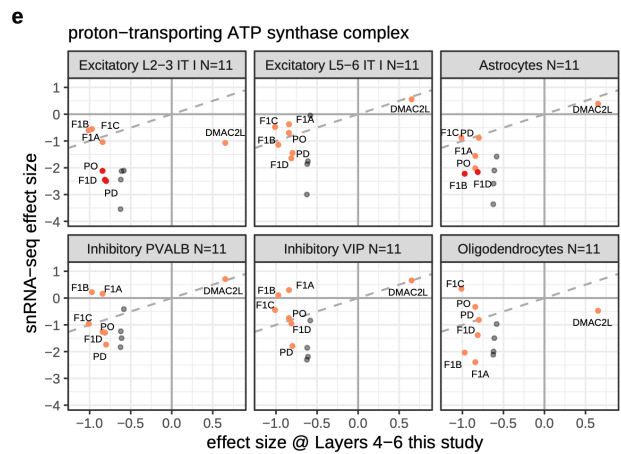
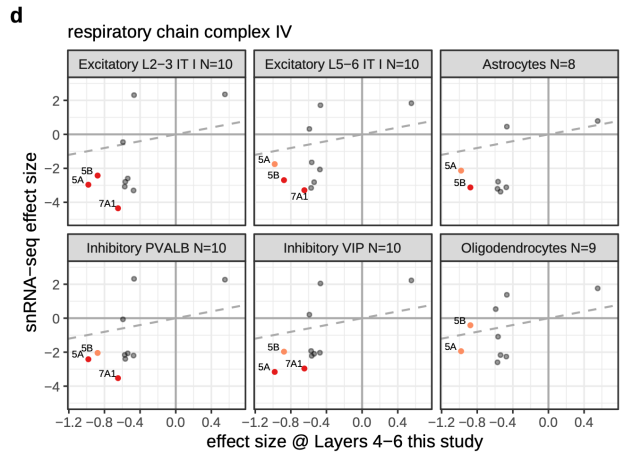
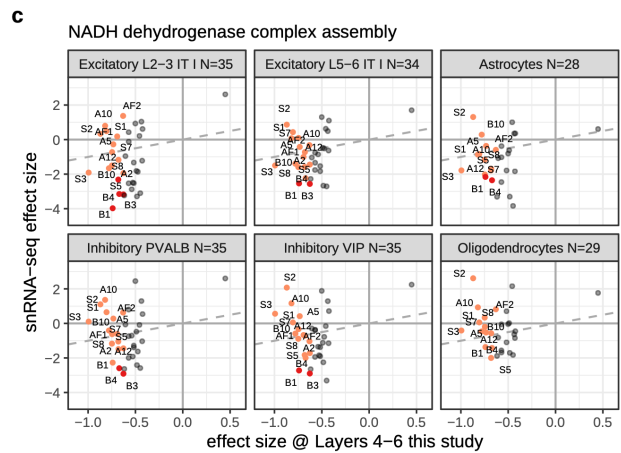
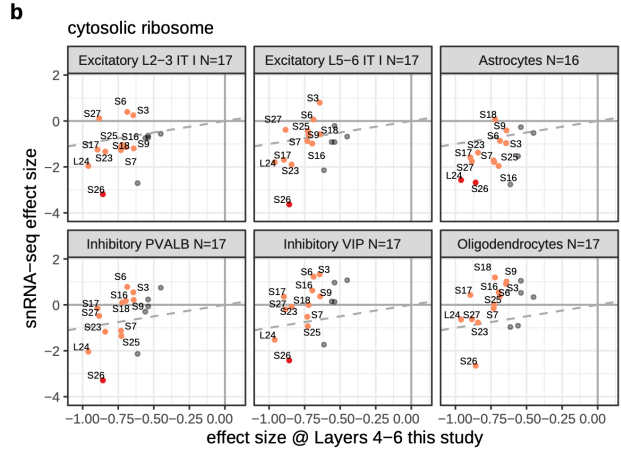
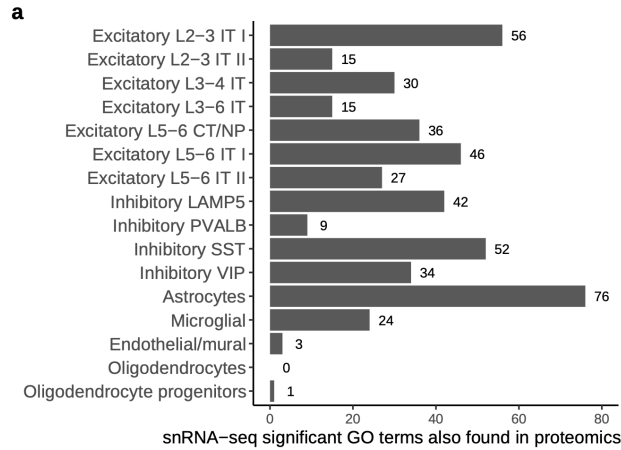


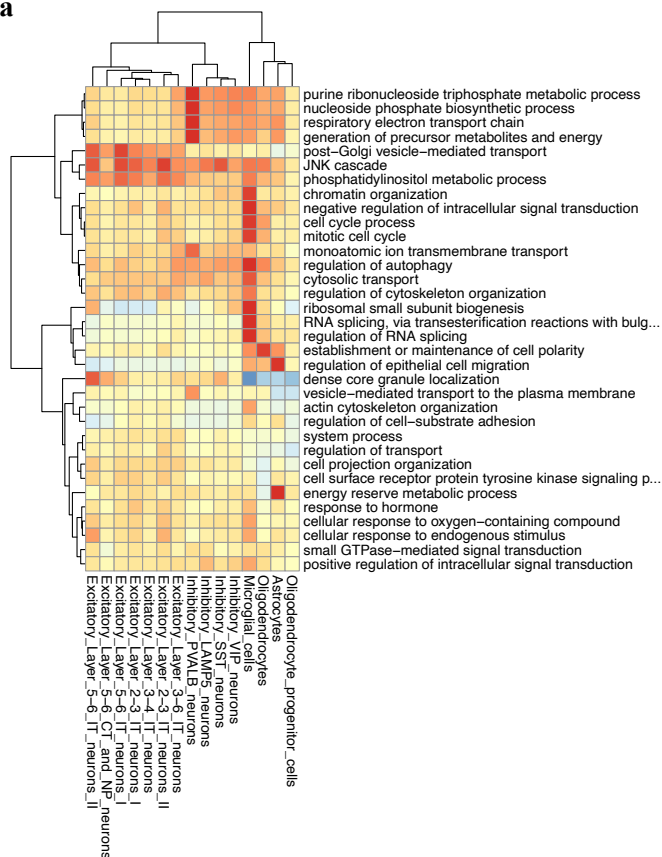
Figure S8. Comparison of schizophrenia (SCZ) associated proteins and pathways between proteomics and snRNA-seq. Variant of main Figure 4, using data from Layers 4-6 for panels b-f.

Comparison between proteomics and snRNA-seq data analogous to main Figure 4, with same layout of panels, but here using proteomics data from Layers 4-6 for panels b-f (whereas main Figure 4 shows data from Layers 1-3).

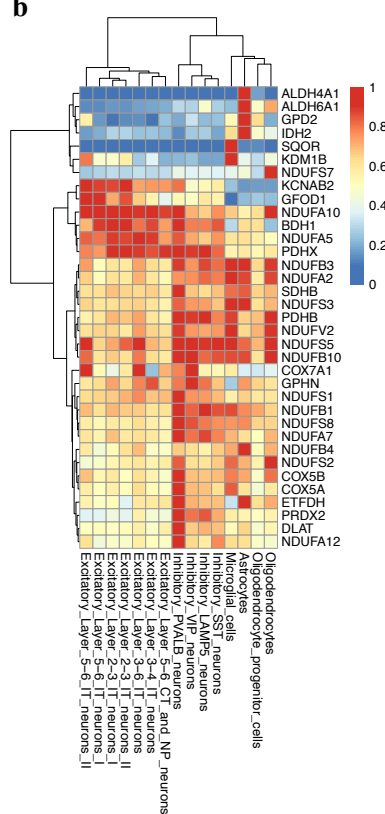
a) For 338 significant GO terms identified across Layer sets in the proteomics dataset (Fig. 3), the overlap with significant GO terms identified in respective cell-types in the accompanying snRNA-seq study was counted. **b-f)** Comparison between protein effect sizes observed in the proteomics control vs SCZ analysis of Layers 1-3 presented in Fig. 1b (Supplementary Data 2) on the x-axis and respective DESeq2 effect sizes observed in the accompanying snRNA-seq study by Bast et al. on the y-axis for 5 selected gene sets (plot title). The number of proteins in each plot is indicated by 'N' in the title. Gene symbols were shortened for visual clarity as follows, **b**; 'RPS' and 'RPL' were replaced by 'S' and 'L', **c**; 'NDUF' was removed from labels, **d**; 'COX' was removed from labels, **e**; 'APT5' was removed from labels.

Panels indicate 6 representative cell-types. Proteins significant in proteomics (FDR adjusted p-value < 0.05) are shown in red if their respective snRNA-seq FDR adjusted p-value < 0.3 and orange otherwise. Proteins not significant in proteomics are shown in grey. Source data are provided as a Source Data file.

a



b



c

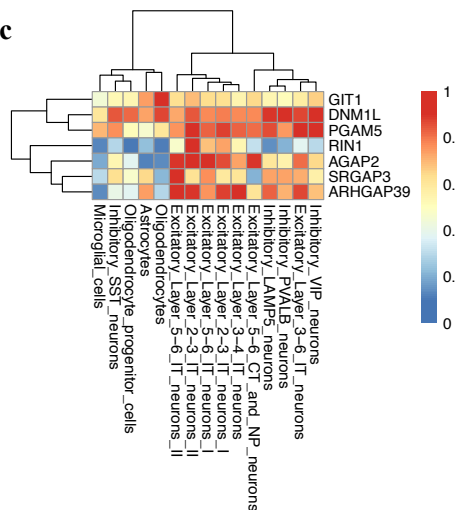


Figure S9. Cell type expression patterns of genes across GO terms enriched in proteomics.

Gene specificity scores were calculated by rescaling gene TPM values from the Bast et al. snRNA-seq dataset by their respective maximum value within each cell type. GO term specificity scores were computed as the (20%) trimmed mean of gene specificity scores from significant genes (in proteomics control vs SCZ, Fig 1b-c, Supplementary Data 2) per GO term. **a)** Significant GO BP terms from control versus SCZ analyses across both Layer sets in the proteomics data (Supplementary Data 5) were reduced to a non-redundant set of GO terms using GOAT (default options) and subselected for those with 3 or more significant genes in the proteomics data (Supplementary Data 2). The resulting 34 GO terms are shown on the y-axis, cell types are shown on x-axis and colors represent the respective GO term specificity scores (based on snRNA-seq expression profiles of gene constituents). **b-c)** To illustrate heterogeneity of expression patterns across cell types at the gene-level, heatmaps of GO terms 'oxidoreductase activity' (**b**) and 'GTPase activator activity' (**c**) are colored by gene specificity scores. The y-axis shows the subset of gene constituents per GO term that were significant in the proteomic analyses of control versus SCZ in either Layer set (Supplementary Data 2). Source data are provided as a Source Data file.

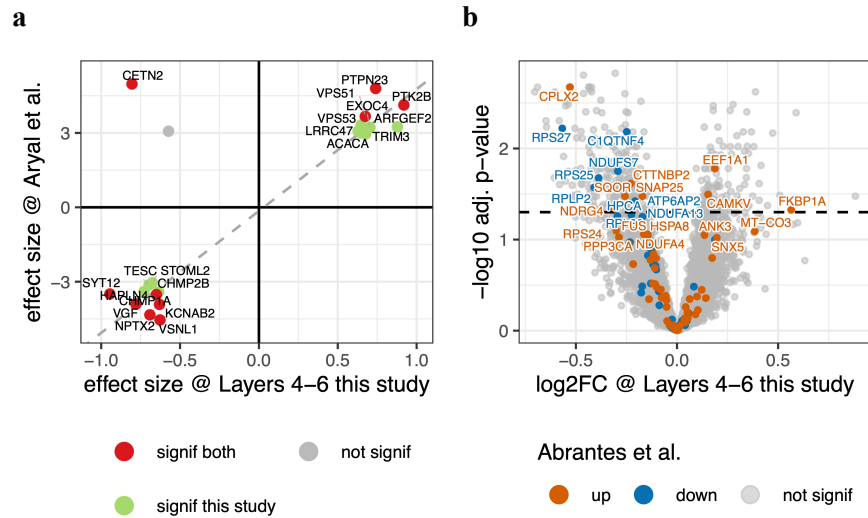


Figure S10. Comparative analysis of data in this study with previously published schizophrenia (SCZ) datasets. Variant of main Figure 5c-d, using data from Layers 4-6.

Panels **a** and **b** are analogous to main Figure 5c-d, respectively, but here using proteomics data from Layers 4-6 (whereas main Figure 5c-d shows data from Layers 1-3).

a) A high degree of similarity in dysregulation is observed for 48 proteins differentially expressed at FDR adjusted p -value < 0.1 in both our study and Aryal et al. The x-axis shows SCZ effect sizes from Layers 4-6 (Fig. 1c). The y-axis shows effect sizes reported by Aryal et al. FDR adjusted p -value < 0.05 was used to classify proteins as significant; red: significant in both studies, green: significant in either study, grey: not significant in either. The dashed line shows a trendline fitted by robust linear regression. **b)** Control vs SCZ analysis of Layers 4-6 presented in Fig. 1c, with proteins color-coded by their regulation after Olanzapine administration in mice as observed with RNA sequencing by Abrantes et al. Up/down-regulated proteins (FDR adjusted p -value < 0.05 in Abrantes et al.) are depicted in orange/blue. Note the anti-correlation of some proteins (e.g. synaptic proteins CPLX2, SNAP25) observed with lower abundance in SCZ patients in this study (x-axis) and the opposite effect observed for Olanzapine treatment in mice by Abrantes et al (color). Source data are provided as a Source Data file.

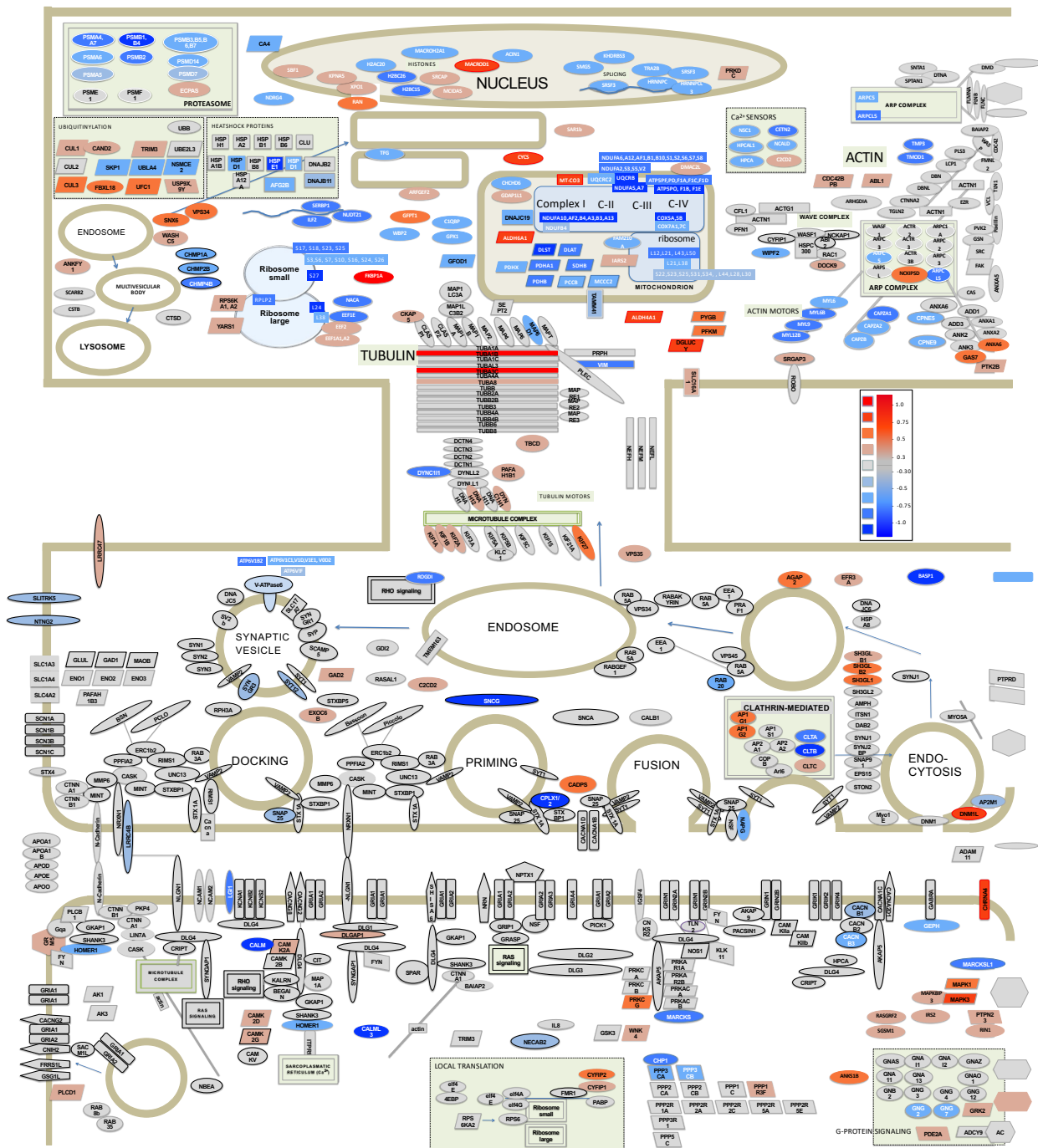


Figure S11. Schematic representation of synaptic subcellular localizations for proteins differentially regulated in schizophrenia.

Differentially regulated proteins in schizophrenia are depicted at their canonical localization. Higher (red) and lower (blue) expression in schizophrenia as indicated by color coding of foldchanges (see legend in middle part of the figure). For the log2 foldchanges shown in this figure, the same limma regression model with DEqMS post-processing was used as with analyses shown in Figure 1b-c, but here using all samples in the dataset and estimating the effect of SCZ across both Layer sets (using a block design in limma, see Methods). Note that most of the dysregulated proteins are not specifically synaptic as becomes evident from a tissue homogenate proteomics approach. Source data are provided as a Source Data file.