

Characterisation of the proteome, diseases and evolution of the human postsynaptic density

Seth G.N. Grant, Alex Bayes, Louie N van de Lagemaat, Mark O Collins,

Mike Dr Croning, Ian R Whittle, Jyoti S Choudhary

▶ To cite this version:

Seth G.N. Grant, Alex Bayes, Louie N van de Lagemaat, Mark O Collins, Mike Dr Croning, et al.. Characterisation of the proteome, diseases and evolution of the human postsynaptic density. Nature Neuroscience, 2010, 10.1038/nn.2719 . hal-00601557

HAL Id: hal-00601557 https://hal.science/hal-00601557

Submitted on 19 Jun 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Characterisation of the proteome, diseases and evolution

of the human postsynaptic density

Àlex Bayés^{1*}, Louie N. van de Lagemaat^{1*}, Mark O. Collins², Mike D.R. Croning¹, Ian R. Whittle³, Jyoti S. Choudhary² & Seth G.N. Grant¹

- 1. Genes to Cognition Programme, Wellcome Trust Sanger Institute, Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK.
- 2. Proteomic Mass Spectrometry. The Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK.
- 3. Division of Clinical Neuroscience, Edinburgh University, Edinburgh, UK.

* These authors contributed equally to this work

Corresponding Author:

Prof. S.G.N. Grant Wellcome Trust Sanger Institute, Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. Tel: 0044 1223 495380, 494908 Fax: 0044 1223 494919 Email: sg3@sanger.ac.uk The postsynaptic density from human neocortex (hPSD) was isolated and 1461 proteins identified. hPSD mutations cause 133 neurological and psychiatric diseases and show enrichment in cognitive, affective and motor phenotypes underpinned by sets of genes. Strong protein sequence conservation within mammalian lineages, particularly in hub proteins, indicates conserved function and organisation in primate and rodent models. The hPSD is a key structure for nervous system disease and behaviour.

Synapses are fundamental structures linking nerve cells and it is essential to characterise human synapse proteins to understand the extent and relevance of synapses to human disease and behavior, and to identify new diagnostic and therapeutic approaches. From neocortical biopsies of nine adults (**Supplementary Table 1**), the hPSD was isolated (**Supplementary Fig. 1a**) and proteomic profiling using LC–MS/MS provided comprehensive protein identification: of a total of 1461 proteins (total hPSD), 748 were detected in all three replicates (consensus hPSD; **Supplementary Table 2**; used for all subsequent analysis, except where explicitly stated otherwise). These data provide a novel resource and are freely available in the G2Cdb database (<u>www.genes2cognition.org/HUMAN–PSD</u>).

This proteomic data allows the first systematic analysis of the disease impact of mutations in hPSD proteins. Annotation of monogenic diseases from the Online Mendelian Inheritance in Man (OMIM) database¹ was chosen for this purpose because it is primarily derived from linkage studies, which are statistically more robust and less prone to artifacts than targeted association studies. These annotations revealed 269 diseases that result from mutations in 199 hPSD genes (14% of all hPSD genes). 133 of these diseases (~50%) are primary nervous system disorders (**Supplementary Tables 3** and **4**). Of these, ~80% were central and ~20% were peripheral nervous system disorders. Similar proportions were seen in the consensus hPSD: 155 diseases arose from mutations in 110 genes and 82 were nervous system diseases. Although the number of brain diseases involved in the hPSD is striking, overall these arise from only 14% of hPSD genes, therefore we expect many more mutations and diseases to be discovered by future human disease genome projects.

To classify these diseases the International Classification of Disease (ICD-10) was used, and from 22 ICD-10 chapters, four represented hPSD diseases: neurology (chapter VI), psychiatry (V), developmental (XVII) and metabolic (IV) diseases (**Fig. 1a**). These included common neurodegenerative diseases (Alzheimer's, Parkinson's, Huntington's), adult and childhood cognitive disorders such as mental retardation, motor disorders such as ataxia or dystonia, epilepsies and many rare diseases.

Each disease class is defined by similar clinical phenotypes and these are systematically categorised in the Human Phenotype Ontology (HPO)². Because HPO phenotypes are linked to gene mutations (via OMIM) we could use gene set enrichment analysis (see **Supplementary Methods**) to identify the phenotypes

most relevant for the hPSD compared to non–PSD brain expressed proteins. As shown in **Figure 1b** and **Supplementary Table 5**, we found the hPSD was significantly enriched in 21 neural phenotypes primarily involving cognition and motor functions. Some were represented by large sets of genes, such as mental retardation (40 genes) and spasticity (20 genes), indicating subsets of the hPSD proteome play key roles in these functions. The enrichment observed in general disease phenotypes (**Fig. 1b**) such as *neurological abnormality* indicates the hPSD is a neuronal structure with a disproportionately high density of neural disease susceptibility.

A powerful way to extend this analysis of hPSD function is to examine the role of hPSD gene orthologs in mice, an important animal model of human disease. Moreover, there is a wider range of behavioural, physiological and anatomical phenotypes that can be investigated for enrichment in the hPSD. From a database analogous to HPO (Mammalian Phenotype Ontology, MPO)³, the hPSD was enriched in 77 neural phenotypes (compared to other sets of brain proteins, **Supplementary Table 6**), which included cognitive and motor phenotypes, consistent with the enrichments found in the HPO analysis (**Fig. 1b** and **Supplementary Table 7**). The MPO also bridges to cellular level data, where we found neuronal morphology (83 genes), neurodegeneration (23 genes), synaptic transmission (38 genes) and plasticity (41 genes) were hPSD enriched phenotypes. The sets of genes underpinning these enriched phenotypes include components of known signalling mechanisms; for example in the *learning/memory/conditioning* gene set (57 genes),

the *N*-methyl–D–aspartate (NMDA) receptor and PSD–95 interacting proteins were present, along with other proteins that are potentially involved with the same signalling process.

We next examined the conservation of hPSD protein coding sequences between humans, other primates (chimpanzee, *Pan troglodytes* and macaque, *Macaca mulatta*) and rodents (mouse, *Mus musculus* and rat, *Rattus norvegicus*) (**Supplementary Table 8**) using the dN/dS ratio; dN measures the rate of amino acid substitution and dS reflects the background rate of neutral DNA change⁴. When comparing human and mouse, which evolved from lineages that diverged from a last common ancestor (LCA) ~90my ago, the median dN/dS for hPSD genes was very significantly (p < 10⁻¹⁴⁸) less than the whole genome (protein coding genes only) (**Supplementary Fig. 2b**). A similar highly significant conservation in hPSD proteins was found in other pairs of compared species: human/chimp (LCA~6 my), human/macaque (LCA~30 my) and mouse/rat (LCA ~20 my)⁵ showing the purifying selection (conservation) was not unique to the human lineage (**Supplementary Fig. 2c-e**).

Since it has been reported that general brain proteins evolved slower than proteins in other tissues^{6–8}, we asked if hPSD conservation was a reflection of broader conservation observed in brain proteins. Unexpectedly, the hPSD was significantly more conserved than all three non–PSD brain expressed gene sets^{9–11} (**Fig. 2a**, **Supplementary Fig. 3** and **Table 9**). The extent of the hPSD conservation was further revealed by comparison with 7 other neuronal subcellular structures (including mitochondria, endoplasmic reticulum and nucleus), which were all less conserved (**Fig. 2b**, **Supplementary Table 10**). Since postsynaptic complexes are rich in protein interactions¹², and a slower rate of evolution has been reported for proteins with many interactions¹³, we constructed an hPSD interaction network (**Supplementary Fig. 4**). Highly interconnected (hub) proteins and proteins in the multiprotein complexes bound to PSD–95¹⁴ showed significantly lower dN/dS than other hPSD proteins (**Fig. 2c, Supplementary Table 11**). These data indicate that hPSD structural organisation plays a role in the conservation of hPSD protein sequence.

This study reveals the human PSD has a high degree of molecular complexity with over 1000 proteins and deepens our understanding of synaptic disease biology by showing that combinations of proteins regulate the phenotypes of over 130 brain diseases. It is possible, and indeed likely, that the proteins identified represent an overall synaptic parts list, with subsets of synapses containing subsets of these proteins. Further sensitive and high-resolution proteomic work will be necessary to expand our knowledge of regional and microscopic synapse heterogeneity. Our data provides a valuable resource and template for investigating human synapse function and suggests new diagnostic and therapeutic approaches.

REFERENCES

- 1. McKusick, V.A. *Am J Hum Genet* **80**, 588–604 (2007).
- 2. Robinson, P.N. et al. *Am J Hum Genet* **83**, 610–615 (2008).
- 3. Smith, C.L., Goldsmith, C.A. & Eppig, J.T. Genome Biol 6, R7 (2005).
- 4. Hurst, L.D. *Trends Genet* **18**, 486 (2002).
- 5. Hedges, S.B., Dudley, J. & Kumar, S. *Bioinformatics (Oxford, England)* **22**, 2971–2972 (2006).
- 6. Wang, H.Y. et al. *PLoS Biol* **5**, e13 (2007).
- 7. Winter, E.E., Goodstadt, L. & Ponting, C.P. *Genome Res* **14**, 54–61 (2004).
- 8. Khaitovich, P. et al. *Science* **309**, 1850–1854 (2005).
- 9. Harris, L.W. et al. *PLoS One* **3**, e3964 (2008).
- 10. Miao, H. et al. *PLoS One* **3**, e2847 (2008).
- 11. Wang, H. et al. *J Proteome Res* **5**, 361–369 (2006).
- 12. Pocklington, A.J., Cumiskey, M., Armstrong, J.D. & Grant, S.G. *Mol Syst Biol* **2**, 2006 0023 (2006).
- 13. Fraser, H.B., Wall, D.P. & Hirsh, A.E. *BMC Evol Biol* **3**, 11 (2003).
- 14. Fernandez, E. et al. *Mol Syst Biol* **5**, 269 (2009).

Fig. 1. hPSD Diseases and Enriched Phenotypes

a. Distribution of hPSD nervous system diseases in 4 ICD–10 chapters (left chart) with Chapter VI expanded (right chart) to show further subclassifications.

b. Representative human and mouse phenotypes enriched in the hPSD.

Categories (bold) of human and mouse phenotypes with number of genes (hPSD genes). Heatmap compares enrichment of these phenotypes in 4 gene sets relative to the enrichment of the hPSD (shown in red). All other gene sets showed lower enrichment (darker colors with black representing no enrichment).

Neuron. human cortical neuron transcriptome⁹; brain, whole mouse brain proteome; astrocyte, human astrocyte transcriptome¹⁰; and human genome.

Hs, human; Mm, mouse.

HPO and MPO phenotype IDs in brackets after each named phenotype.

Fig. 2. hPSD sequence conservation

a. Cumulative frequency plot of dN/dS values for hPSD and non–PSD genes expressed in cortical neurons⁹, and human genome. hPSD neuronal genes are more constrained than non–PSD neuronal genes ($p < 10^{-11}$).

b. Median mouse-human dN/dS shown for hPSD and subcellular structures expressed in cortical neurons⁹. Significance: p < 0.05 (*), p < 0.001 (***).

c, Boxplots of dN/dS distribution in hPSD hub proteins (>15 interactions, n = 23), non-hubs (\leq 15 interactions, n = 725) and the tandem affinity purification of PSD-95 complex¹⁴.

Acknowledgements

We thank C.P. Ponting, P.J. Brophy. R.A.W. Frank, N.H. Komiyama, M.V. Kopanitsa, T. J. Ryan and members of the Genes to Cognition Programme for critical comments on the manuscript and discussions. AB is supported by EMBO and the European Commission. LVL, MOC, MDRC, JSC and SGNG are supported by the Wellcome Trust. IRW is supported by Scottish Higher Education Funding Council and by grants from MRC, NIH and Melville Trust. This study was approved by Lothian Region Ethics Committee /2004/4/16, and informed consent was obtained from all donors. We thank the tissue donors, without whom this study would have been impossible.

Author Contributions

IRW provided brain samples; AB, MOC, JSC performed proteomic analysis. AB performed OMIM, ICD–10 and evolutionary analyses, LVL performed network and enrichment analyses. LVL, MDRC, AB and SGNG wrote the manuscript.

Author Information

No competing interests

Supplementary Table Legends

Supplementary Table 1: Information about Biological Samples.

Patient code, age, sex and reason for surgery are shown. Also included is the precise cortical location of each biopsy and in what hPSD fractionation it was used.

Supplementary Table 2: Protein Identifications and Proteomic Data.

For each identified protein several identification (ID) numbers from biological databases are given: G2Cdb ID, International Protein Index (IPI) ID, Approved Gene Name from HUGO Genome Nomenclature Committee (HGNC), HCNC ID, Ensembl ID, Entrez Gene ID, OMIM (Gene) ID and UniProt ID. The number of total and uniquely identified peptides for each replicate is also provided. Proteins found with two or more peptides in all replicates are classified as members of the consensus hPSD.

Supplementary Table 3: Summary of OMIM Diseases.

Diseases identified in the Total hPSD or the Consensus hPSD were classified according to the system they affected: non Nervous System Disease, Peripheral Nervous System Diseases, Central Nervous System Diseases or affecting both the Peripheral and Central Nervous Systems. Diseases manifested in the Central Nervous System (the sum of central only and peripheral and central) were classified in ICD–10 chapters. 'Diseases of the Nervous System' were further classified into their sub–chapters. Finally, the number of proteins causing diseases of each group is shown.

Supplementary Table 4: OMIM Diseases Identified Among Total hPSD Genes.

For each protein causing an OMIM disease the following information is given: G2Cdb ID, Approved (HGNC) Gene Symbol. Also shown is whether it is present in the consensus PSD (found in all three proteomic replicates), if it causes a Peripheral Nervous System Disease or a Central Nervous System Disease or a Peripheral and Central Nervous System Disease or a Non Nervous system Disease. In addition, OMIM Disease ID, OMIM Disease Description, Disease age of onset, ICD–10 Category for CNS Disease. and ICD–10 Sub–Category for CNS Diseases. Neoplasms and other diseases caused by somatic mutations have been excluded from the list.

Supplementary Table 5: Human Neural Phenotype gene set enrichment analysis. All phenotypes of neural origin from Human Phenotype Ontology (HPO) are included. For each HPO phenotype the HPO ID and description are given. The number of genes from the genome or the hPSD causing each phenotype is shown under the Genome and hPSD columns. Phenotype overrepresentation in the hPSD is with respect to the genome, and approved (HGNC) symbols for relevant hPSD genes are given. For each background dataset (mouse brain proteome, human cortical neuron transcriptome, and human astrocyte), number of phenotype genes present in the non-hPSD portion of the background dataset are listed in the 'dataset' column. Overrepresentation of phenotype genes in the background sets is computed relative to the genome in the column 'Overrep (dataset/ genome)'. P-value of overrepresentation of the phenotype in the hPSD relative to the background set is denoted 'P-value (hPSD > Other)' and is computed based on the observed overrepresentation in the hPSD and compared to the overrepresentation that would be observed if the hPSD were only as overrepresented in the particular phenotype as the background set. P-values were calculated using binomial statistics and corrected using the Benjamini–Hochberg false discovery rate procedure at $\alpha = 0.05$.

Supplementary Table 6: Brain datasets used in phenotype enrichment analysis. For each of the three sets of brain molecules (mouse brain proteome, human cortical neuron transcriptome and human astrocyte transcriptome) all Ensembl identifiers are given.

Supplementary Table 7: Mammalian Neural Phenotype gene set enrichment Analysis.

All phenotypes of neural origin from Mammalian Phenotype Ontology (MPO) are included, as well as two manually defined phenotypes representing synaptic transmission and synaptic plasticity. For each MPO phenotype the MPO ID and description are given. The number of genes from the genome or the hPSD causing each phenotype is shown under the Genome and hPSD columns. Phenotype overrepresentation in the hPSD is with respect to the genome, and approved (HGNC) symbols for relevant hPSD genes are given. For each background dataset (mouse brain proteome, human cortical neuron transcriptome, and human astrocyte

13

transcriptome), number of phenotype genes present in the non–hPSD portion of the background dataset are listed in the 'dataset' column. Overrepresentation of phenotype genes in the background sets is computed relative to the genome in the column 'Overrep (dataset/ genome)'. P–value of overrepresentation of the phenotype in the hPSD relative to the background set is denoted 'P–value (hPSD > Other)' and is computed based on the observed overrepresentation in the hPSD and compared to the overrepresentation that would be observed if the hPSD were only as overrepresented in the particular phenotype as the background set. P–values were calculated using binomial statistics and corrected using the Benjamini–Hochberg false discovery rate procedure at $\alpha = 0.05$.

Supplementary Table 8: Comparison of dN/dS between Genome and hPSD.

dN, dS and dN/dS values obtained from Ensembl 53 for Human–Mouse, Human– Macaque, Human–Chimpanzee and Mouse–Rat.Genome as well as total and Consenus hPSD values are shown. Median values for dN, dS and dN/dS are also shown.

Supplementary Table 9: Mouse and Human Brain Datasets dN/dS Analysis.

Genes from the human genome, mouse brain proteome, human cortical neuron transcriptome and human cortex proteome for which there was dN/dS information were divided into those included in the hPSD and those absent from it. The cumulative dN/dS distributions are computed for each dataset.

14

Supplementary Table 10: dN/dS Values for genes expressed in human neurons classified by cellular component.

Mouse to human dN/dS values from Ensembl 53 for genes expressed in human cortical neurons classified by cellular component.

Supplementary Table 11: Analysis of dN/dS on Hub, non–Hub and TAP–PSD–95 proteins.

Mouse to human dN, dS and dN/dS values obtained from Ensembl 53 are shown for Hub consensus hPSD proteins (> 15 interactions), Non–Hub consensus hPSD proteins (<= 15 interactions) as well as for proteins isolated using tandem affinity purification on PSD–95.

Supplementary Table 12: Mouse to Human dN/dS Values in hPSD and Other Organelle Proteomes.

Mouse to human dN/dS values from Ensembl 53 for the following proteomics sets are provided: whole genome, hPSD, consensus hPSD, mitochondrion, endoplasmic reticulum (ER), Golgi apparatus, ER Golgi vesicle, early endosome, recycling endosome, plasma membrane, proteasome, cytosol, and nucleus.

Supplementary Figure Legends

Supplementary Figure 1. Isolation of Postsynaptic Densities from Human Neocortex.

a, Location of the cortical samples obtained by biopsy used in this analysis.

15

b, Protein staining of SDS–PAGE of PSD preparations.

c, Immunoblotting of the synaptosomal (S#) and PSD (P#) fractions from each preparation with known postsynaptic markers. Enrichment in the hPSD fraction was observed for all markers: Neurotransmitter receptors; NMDA receptor subunit 1 (NR1), NMDA receptor subunit 2A (NR2A), NMDA receptor subunit 2B (NR2B); Scaffold proteins; postsynaptic density protein 95 (PSD–95), postsynaptic density protein 93 (PSD93), Synapse–associated protein 102 (SAP102), Synapse–associated protein 97 (SAP97); Signaling proteins; Calcium/calmodulin–dependent protein kinase type II alpha chain (CAMKIIα), Ras GTPase–activating protein SynGAP (SYNGAP), Insulin receptor substrate 1 (IRS1), Nitric oxide synthase, brain specific (nNOS) and Ras–related C3 botulinum toxin substrate 1 (RAC1).

d, Eight proteins from the list of genes expressed in human cortical neurons¹ but absent from the total hPSD set were immunoblotted for the following cellular fractions: Whole homogenate (Homo), Nuclear Fraction (Nucl), Cytosolic fraction (Cyto) and postsynaptic fraction (hPSD). Note all proteins are absent from the hPSD fraction.

Supplementary Figure 2. Comparative Analysis of Sequence Conservation of hPSD genes.

a, Phylogenetic tree scaled to time based on molecular data. Last Common Ancestor (LCA) indicated by branch points. Primate and rodent species analysed are represented. b-e, Median dN/dS of genome, total hPSD and consensus hPSD for pairs of species (indicated above graph). p values between genome and tPSD or cPSD indicated in each comparison.

Supplementary Figure 3. hPSD constraint relative to other brain proteins

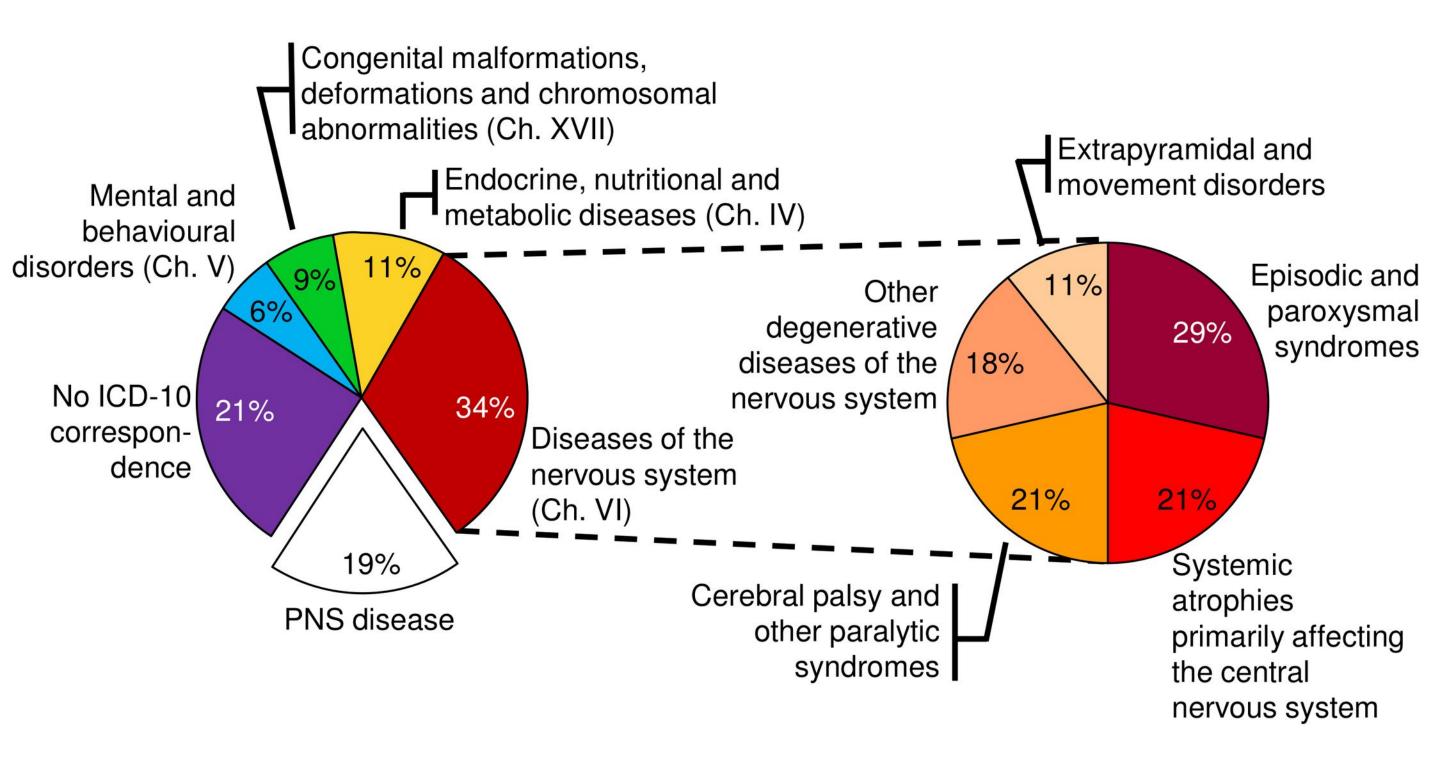
a–d. hPSD is constrained relative to other brain proteins. Cumulative frequency plots of dN/dS distributions for human PSD and non–hPSD brain subsets and whole genome. The hPSD was more highly constrained than non–hPSD brain proteins measured using 4 datasets: a, human cortex proteome (p < 10^{-6}); b, human cortical neuron transcriptome (p < 10^{-9}); c, mouse brain proteome (p < 10^{-29}) and d, mouse brain proteome of transmembrane proteins (p < 10^{-11}).

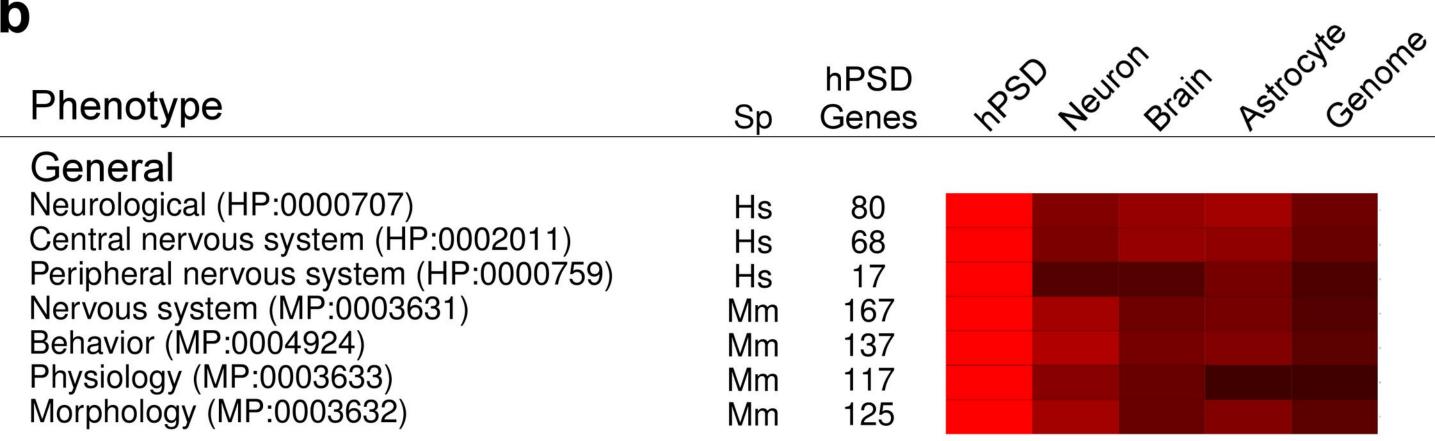
e,f. hPSD is constrained relative to other subcellular structures.

Median mouse–human dN/dS shown for hPSD and other subcellular structures profiled by (e) transcriptome from human neocortical neurons and (f) mass spectrometry based proteomics in liver. All structures showed a median dN/dS above the hPSD: (e) nucleus p< 10^{-20} ; plasma membrane p = $< 10^{-14}$; golgi p $< 10^{-6}$; cytosol p $< 10^{-10}$; ER p $< 10^{-17}$; Early Endosome p = 0.0052; and Mitochondria p $< 10^{-36}$). (f) nucleus p= 0.24; recycling endosome p= 0.028; plasma membrane p = 0.0057; golgi p $< 10^{-4}$; cytosol p $< 10^{-11}$; ER Golgi Vesicle p $< 10^{-4}$; Early Endosome p = 0.0059; ER p $< 10^{-7}$ and Mitochondria p $< 10^{-29}$). Significance: p< 0.05 (*), p< 0.01 (***).

Supplementary Figure 4. hPSD protein interaction network

Proteins in the network are coloured by their dN/dS ratio. dN/dS ranges from 0.0 (no non–synonymous amino acid changes) to 0.048 (median dN/dS in the PSD) to 0.117 (genomic median). HGNC symbols of hub proteins are shown.





Hs

Mm

Mm

Mm

Mm

Hs

Hs

Hs

Hs

Hs

Hs

Mm

Mm

Mm

Mm

Mm

Mm

Mm

Peripheral nervous system (HP:0000759) Nervous system (MP:0003631) Behavior (MP:0004924) Physiology (MP:0003633) Morphology (MP:0003632)	
Cognitive/affect Mental retardation (HP:0001249) Learning/memory/conditioning (MP:000206	6

63) Emotion/affect behavior (MP:0002572) Social/conspecific interaction (MP:0002557) Addiction/drug abuse (MP:0002568)

Motor

Spasticity (HP:0001257)
Muscle weakness (HP:0001324)
Nystagmus (HP:0000639)
Dystonia (HP:0001332)
Gait disturbance (HP:0001288)
Diminished movement (HP:0002374)
Motor capabilities (MP:0002066)

Physiology

Synaptic transmission Synaptic plasticity Seizures (MP:0002064) Neuron physiology (MP:0004811) Touch/ nociception (MP:0001968) Induced seizures (MP:0000950)

Morphology

Myelination (HP:0002520) Brain morphology (MP:0002152) Neuron morphology (MP:0002882) Somatic NS morphology (MP:0002752) Neurodegeneration (MP:0002229) Spinal cord morphology (MP:0000955) Myelination (MP:0000920)

Hs	17
Λm	167
Λm	137
Лm	117
٨m	125

40

57

38

21

12

20

19

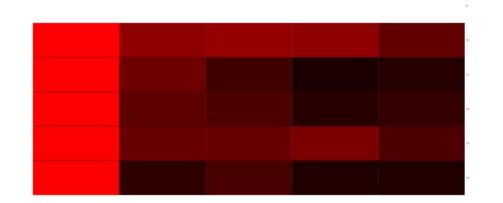
17

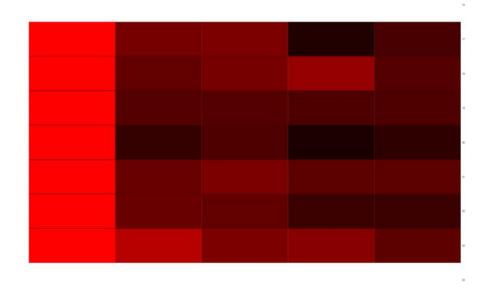
13

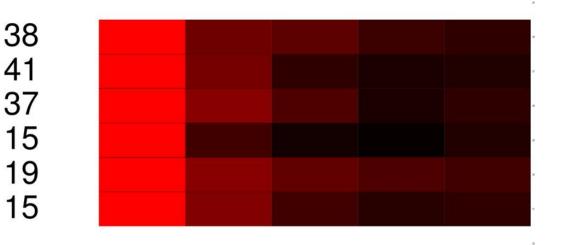
12

11

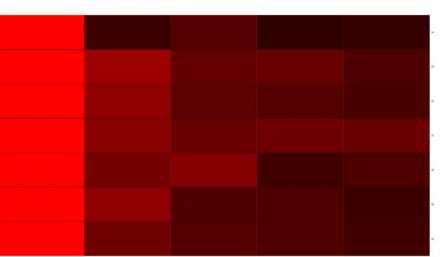
91







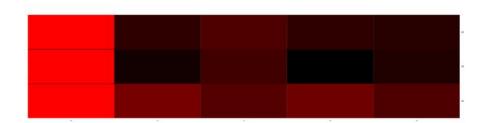


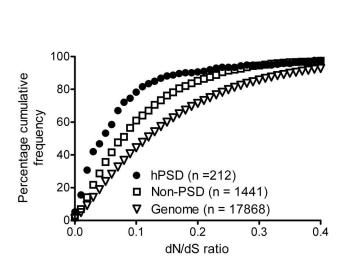


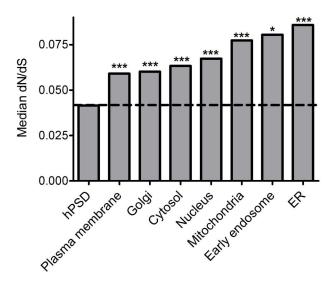
Other

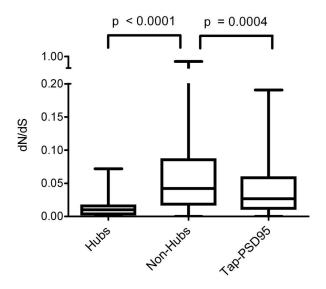
Neurological lab. findings (HP:0003129) CSF findings (HP:0002921) Peripheral neuropathy (HP:0009830)











b

С