

Published in final edited form as:

*Nat Neurosci.* 2011 January ; 14(1): 19–21. doi:10.1038/nn.2719.

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

Àlex Bayés<sup>1,\*</sup>, Louie N. van de Lagemaat<sup>1,\*</sup>, Mark O. Collins<sup>2</sup>, Mike D.R. Croning<sup>1</sup>, Ian R. Whittle<sup>3</sup>, Jyoti S. Choudhary<sup>2</sup>, and Seth G.N. Grant<sup>1</sup>

<sup>1</sup>Genes to Cognition Programme, Wellcome Trust Sanger Institute, Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK

<sup>2</sup>Proteomic Mass Spectrometry. The Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK

<sup>3</sup>Division of Clinical Neuroscience, Edinburgh University, Edinburgh, UK

### Abstract

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 ([www.genes2cognition.org/HUMAN-PSD](http://www.genes2cognition.org/HUMAN-PSD)).

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) database<sup>1</sup> 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

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, sg3@sanger.ac.uk.

\*These authors contributed equally to this work

#### 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.

No competing interests

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)<sup>2</sup>. 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)<sup>3</sup>, 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 change<sup>4</sup>. 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)<sup>5</sup> 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<sup>6-8</sup>, 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 interactions<sup>12</sup>, and a slower rate of evolution has been reported for proteins with many interactions<sup>13</sup>, 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.

## Supplementary Material

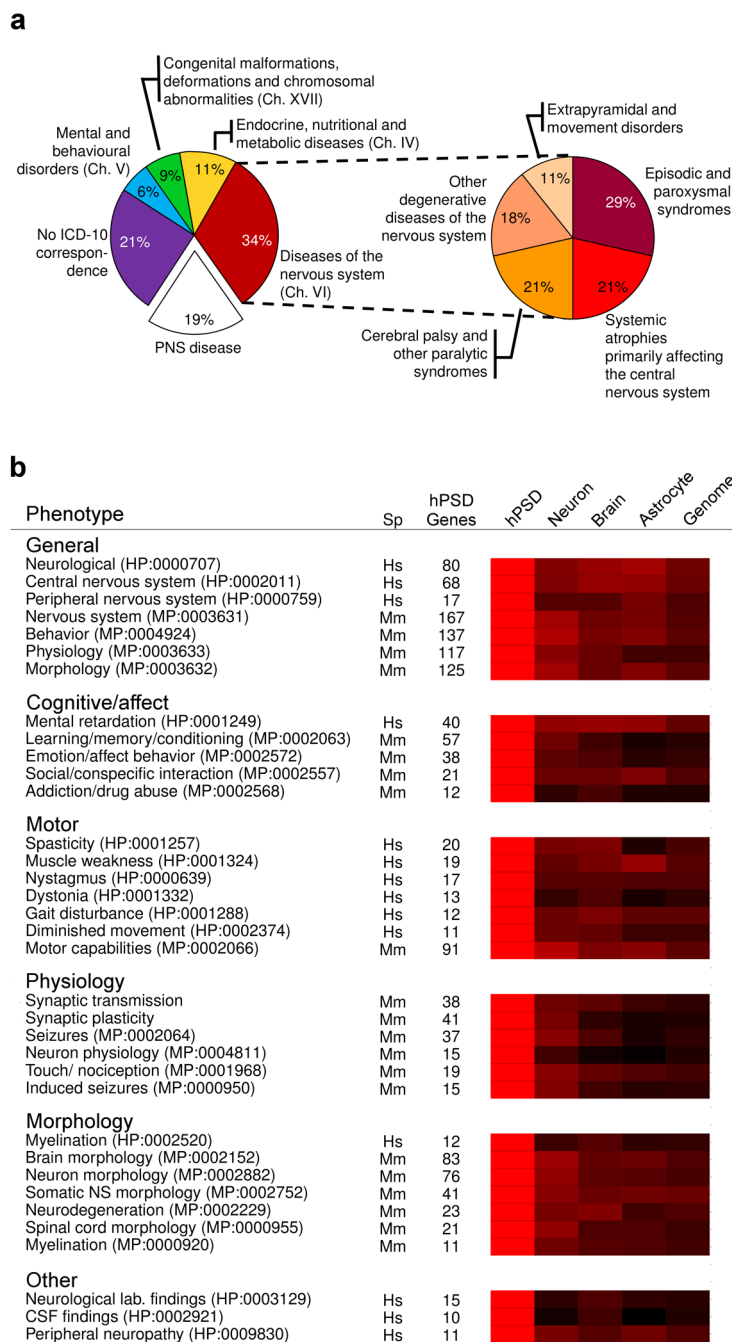
Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

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.

## REFERENCES

1. McKusick VA. *Am J Hum Genet.* 2007; 80:588–604. [PubMed: 17357067]
2. Robinson PN, et al. *Am J Hum Genet.* 2008; 83:610–615. [PubMed: 18950739]
3. Smith CL, Goldsmith CA, Eppig JT. *Genome Biol.* 2005; 6:R7. [PubMed: 15642099]
4. Hurst LD. *Trends Genet.* 2002; 18:486. [PubMed: 12175810]
5. Hedges SB, Dudley J, Kumar S. *Bioinformatics (Oxford, England).* 2006; 22:2971–2972.
6. Wang HY, et al. *PLoS Biol.* 2007; 5:e13. [PubMed: 17194215]
7. Winter EE, Goodstadt L, Ponting CP. *Genome Res.* 2004; 14:54–61. [PubMed: 14707169]
8. Khaitovich P, et al. *Science.* 2005; 309:1850–1854. [PubMed: 16141373]
9. Harris LW, et al. *PLoS One.* 2008; 3:e3964. [PubMed: 19088852]
10. Miao H, et al. *PLoS One.* 2008; 3:e2847. [PubMed: 18716680]
11. Wang H, et al. *J Proteome Res.* 2006; 5:361–369. [PubMed: 16457602]
12. Pocklington AJ, Cumiskey M, Armstrong JD, Grant SG. *Mol Syst Biol.* 2006; 2:0023. 2006. [PubMed: 16738568]
13. Fraser HB, Wall DP, Hirsh AE. *BMC Evol Biol.* 2003; 3:11. [PubMed: 12769820]
14. Fernandez E, et al. *Mol Syst Biol.* 2009; 5:269. [PubMed: 19455133]



**Fig. 1. hPSD Diseases and Enriched Phenotypes**

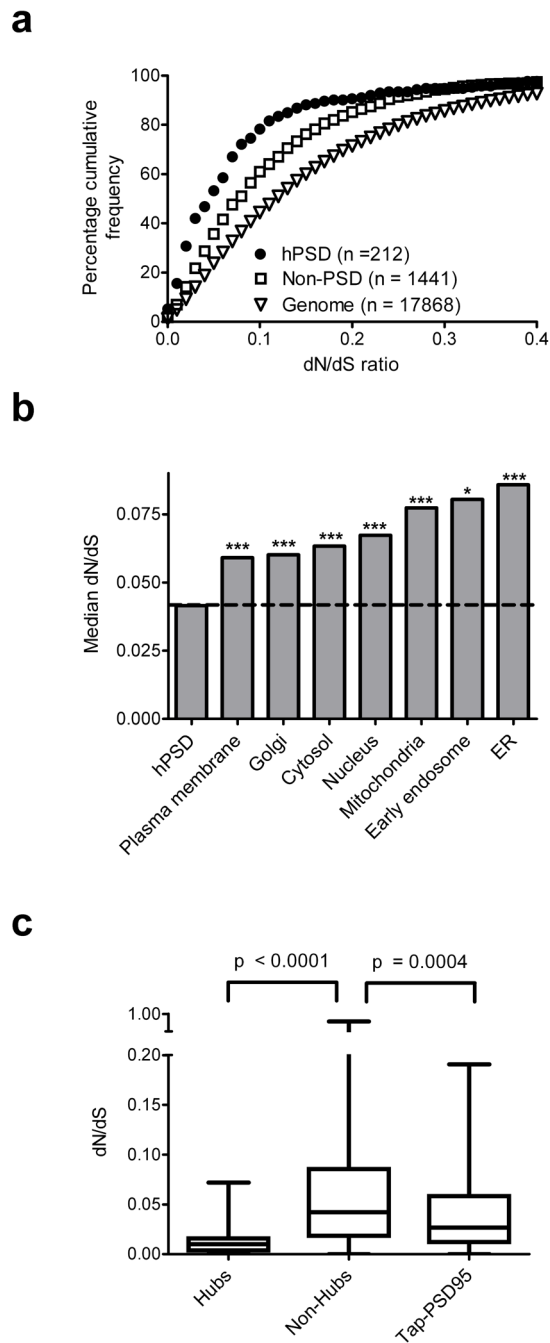
a. Distribution of hPSD nervous system diseases in 4 ICD–10 chapters (left chart) with Chapter VI expanded (right chart) to show further subclassifications.

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 neurons<sup>9</sup>, 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 neurons<sup>9</sup>. Significance:  $p < 0.05$  (\*),  $p < 0.001$  (\*\*\*)

**c.** Boxplots of dN/dS distribution in hPSD hub proteins (>15 interactions,  $n = 23$ ), non-hubs (15 interactions,  $n = 725$ ) and the tandem affinity purification of PSD-95 complex<sup>14</sup>.