nature neuroscience

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

Àlex Bayés[1,4], Louie N van de Lagemaat[1,4], Mark O Collins[2], Mike D R Croning[1], Ian R Whittle[3], Jyoti S Choudhary[2] & Seth G N Grant[1]

We isolated the postsynaptic density from human neocortex (hPSD) and identified 1,461 proteins. hPSD mutations cause 133 neurological and psychiatric diseases and were enriched in cognitive, affective and motor phenotypes underpinned by sets of genes. Strong protein sequence conservation in mammalian lineages, particularly in hub proteins, indicates conserved function and organization in primate and rodent models. The hPSD is an important structure for nervous system disease and behavior.

Synapses are fundamental structures linking nerve cells and it is essential to characterize human synapse proteins to understand the extent and relevance of synapses to human disease and behavior and to identify new diagnostic and therapeutic approaches. We isolated hPSD from neocortical biopsies of nine adults (**Supplementary Table 1** and **Supplementary Fig. 1a**) and performed proteomic profiling using LC-MS/MS to comprehensively identify a total of 1,461 proteins (total hPSD), 748 of which were detected in all three replicates (consensus hPSD, used for all subsequent analysis, except where explicitly stated otherwise; **Supplementary Table 2**). These data are a freely available resource in the G2Cdb database (http://www.genes2cognition.org/HUMAN-PSD). This study was approved by Lothian Region Ethics Committee/2004/4/16, and informed consent was obtained from all donors.

This proteomic data allows, to the best of our knowledge, the first systematic analysis of the effects of mutations in hPSD proteins on disease. Annotation of monogenic diseases from the Online Mendelian Inheritance in Man (OMIM) database[1] 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). Of these diseases, 133 (~50%) are primary nervous system disorders (**Supplementary Tables 3** and **4**); ~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, 82 of which were nervous system diseases. Although the number of brain diseases in which the hPSD is involved is notable, these arise from only 14% of hPSD genes, suggesting that there are many more mutations and diseases to be discovered by future human disease genome projects.

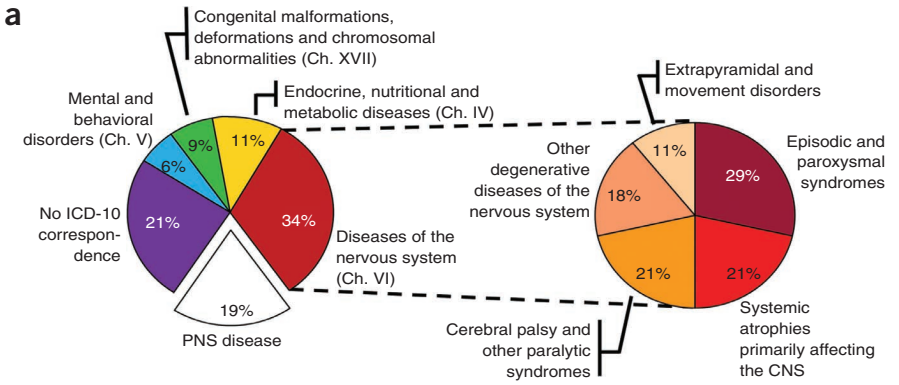
To classify these diseases, we used the International Classification of Disease (ICD-10), and from 22 ICD-10 chapters, four represented hPSD diseases: neurology (chapter VI), psychiatry (V), developmental (XVII) and metabolic (IV) diseases (**Fig. 1a**). These included common neurodegenerative diseases (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 clinical phenotypes and these are systematically categorized in the Human Phenotype Ontology (HPO)[2]. Because HPO phenotypes are linked to gene mutations (via OMIM) we were able to use gene set enrichment analysis (**Supplementary Methods**) to identify the phenotypes most relevant for the hPSD compared with brain proteins that are not expressed in PSD. We found the hPSD was significantly enriched in 21 neural phenotypes ($P < 0.05$) that primarily involve cognition and motor functions (**Fig. 1b** and **Supplementary Table 5**). Some were represented by large sets of genes, such as mental retardation (40 genes) and spasticity (20 genes), indicating that subsets of the hPSD proteome are important for these functions. The enrichment observed in general disease phenotypes (**Fig. 1b**), such as neurological abnormality, indicates that 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 mouse, an important animal model of human disease. Moreover, there is a wider range of behavioral, physiological and anatomical phenotypes that can be investigated for enrichment in the hPSD. From a database analogous to HPO (Mammalian Phenotype Ontology, MPO)[3], we found that the hPSD was enriched in 77 neural phenotypes (compared with 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 that 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 signaling mechanisms; for example, in the learning/memory/conditioning gene set (57 genes), the NMDA receptor and PSD-95–interacting proteins were present, along with other proteins that are potentially involved in the same signaling process.

[1]Genes to Cognition Programme, Wellcome Trust Sanger Institute, Genome Campus, Hinxton, Cambridgeshire, UK. [2]Proteomic Mass Spectrometry, Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, UK. [3]Division of Clinical Neuroscience, Edinburgh University, Edinburgh, UK. [4]These authors contributed equally to this work. Correspondence should be addressed to S.G.N.G. (sg3@sanger.ac.uk).

Received 23 June; accepted 12 November; published online 19 December 2010; doi:10.1038/nn.2719



b

Phenotype	Sp	hPSD genes
General		
Neurological (HP:0000707)	Hs	80
CNS (HP:0002011)	Hs	68
Peripheral nervous system (HP:0000759)	Hs	17
Nervous system (MP:0003631)	Mm	167
Behavior (MP:0004924)	Mm	137
Physiology (MP:0003633)	Mm	117
Morphology (MP:0003632)	Mm	125
Cognitive/affect		
Mental retardation (HP:0001249)	Hs	40
Learning/memory/conditioning (MP:0002063)	Mm	57
Emotion/affect behavior (MP:0002572)	Mm	38
Social/conspecific interaction (MP:0002557)	Mm	21
Addiction/drug abuse (MP:0002568)	Mm	12
Motor		
Spasticity (HP:0001257)	Hs	20
Muscle weakness (HP:0001324)	Hs	19
Nystagmus (HP:0000639)	Hs	17
Dystonia (HP:0001332)	Hs	13
Gait disturbance (HP:0001288)	Hs	12
Diminished movement (HP:0002374)	Hs	11
Motor capabilities (MP:0002066)	Mm	91
Physiology		
Synaptic transmission	Mm	38
Synaptic plasticity	Mm	41
Seizures (MP:0002064)	Mm	37
Neuron physiology (MP:0004811)	Mm	15
Touch/nociception (MP:0001968)	Mm	19
Induced seizures (MP:0000950)	Mm	15
Morphology		
Myelination (HP:0002520)	Hs	12
Brain morphology (MP:0002152)	Mm	83
Neuron morphology (MP:0002882)	Mm	76
Somatic NS morphology (MP:0002752)	Mm	41
Neurodegeneration (MP:0002229)	Mm	23
Spinal cord morphology (MP:0000955)	Mm	21
Myelination (MP:0000920)	Mm	11
Other		
Neurological lab findings (HP:0003129)	Hs	15
CSF findings (HP:0002921)	Hs	10
Peripheral neuropathy (HP:0009830)	Hs	11

Figure 1 hPSD diseases and enriched phenotypes. (**a**) Distribution of hPSD nervous system diseases in four ICD-10 chapters (left) with chapter VI expanded (right) to show further subclassifications. (**b**) Representative human and mouse phenotypes enriched in the hPSD. Categories (bold) of human and mouse phenotypes with numbers of genes (hPSD genes) are shown. Heat map comparing enrichment of these phenotypes in four 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). Astrocyte, human astrocyte transcriptome[12]; brain, whole mouse brain proteome[13]; genome, human genome; Hs, human; Mm, mouse; neuron, human cortical neuron transcriptome[14]. HPO and MPO phenotype IDs are shown in brackets after each named phenotype.

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[4]. When comparing human and mouse, which evolved from lineages that diverged from a last common ancestor (LCA) ~90 million years ago, the median dN/dS for (total) hPSD genes was very significantly ($P < 10^{-148}$) less than that of 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 million years ago $P < 10^{-87}$), human/macaque (LCA ~30 million years ago $P < 10^{-71}$) and mouse/rat (LCA ~20 million years ago $P < 10^{-77}$)[5], indicating that the purifying selection (conservation) was not unique to the human lineage (**Supplementary Fig. 2c–e**).

As it has been reported that general brain proteins evolve slower than proteins in other tissues[6–8], we asked whether hPSD conservation was a reflection of broader conservation observed in brain proteins. Unexpectedly, the hPSD was substantially more conserved than all four non-PSD brain-expressed gene sets[2] (**Fig. 2a**, **Supplementary Fig. 3a–d** and **Supplementary Table 9**). The extent of the hPSD conservation was further revealed by comparison with seven other neuronal subcellular structures (including mitochondria, endoplasmic reticulum and nucleus), which were all less conserved (**Fig. 2b** and **Supplementary Table 10**). As postsynaptic complexes are rich in protein interactions[9] and a slower rate of evolution has been reported for proteins with many interactions[10], we constructed an hPSD interaction network (**Supplementary Fig. 4**). Highly interconnected (hub) proteins and proteins in the multiprotein complexes bound to PSD-95 (ref. 11) showed significantly lower dN/dS than other hPSD proteins (**Fig. 2c** and **Supplementary Table 11**). These data indicate that hPSD structural organization is involved in the conservation of hPSD protein sequence.

Our results indicate that the human PSD has a high degree of molecular complexity, with over 1,000 proteins, and deepen 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 provide a

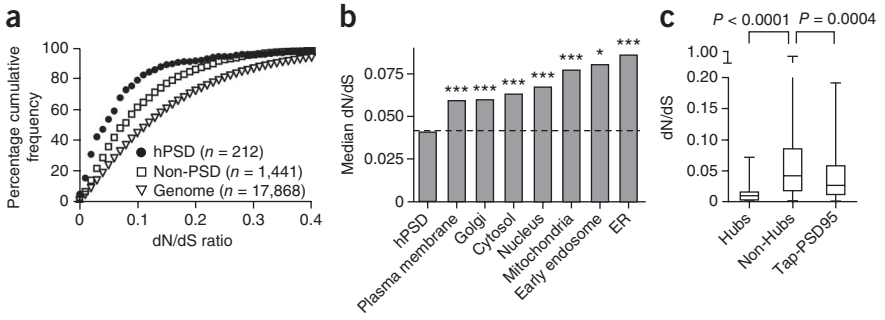


Figure 2 hPSD sequence conservation. (**a**) Cumulative frequency plot of dN/dS values for hPSD and non-PSD genes expressed in cortical neurons[14], and human genome. hPSD neuronal genes are more constrained than non-PSD neuronal genes ($P < 10^{-11}$). (**b**) Median mouse-human dN/dS shown for hPSD and subcellular structures expressed in cortical neurons[14]. $^{*}P < 0.05$, $^{***}P < 0.001$. (See **Supplementary Table 10** and **Supplementary Table 12** for a comparison with proteomically derived organelles). (**c**) Box plots 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[11].

valuable resource and template for investigating human synapse function and suggest new diagnostic and therapeutic approaches.

Note: Supplementary information is available on the Nature Neuroscience website.

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 for discussions. We thank the tissue donors, without whom this study would have been impossible. A.B. is supported by EMBO and the European Commission. L.N.v.L., M.O.C., M.D.R.C., J.S.C. and S.G.N.G. are supported by the Wellcome Trust. I.R.W. is supported by Scottish Higher Education Funding Council and by grants from the Medical Research Council, the US National Institutes of Health and the Melville Trust.

AUTHOR CONTRIBUTIONS

I.R.W. provided brain samples. A.B., M.O.C. and J.S.C. performed proteomic analysis. A.B. performed OMIM, ICD-10 and evolutionary analyses. L.N.v.L. performed network and enrichment analyses. M.D.R.C. integrated data into G2Cdb. L.N.v.L., M.D.R.C., M.O.C., A.B. and S.G.N.G. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/natureneuroscience/.
Reprints and permissions information is available online at http://www.nature.com/reprintsandpermissions/.

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* **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. Pocklington, A.J., Cumiskey, M., Armstrong, J.D. & Grant, S.G. *Mol. Syst. Biol.* **2**, 2006. 0023 (2006).
10 Fraser, H.B., Wall, D.P. & Hirsh, A.E. *BMC Evol. Biol.* **3**, 11 (2003).
11. Fernández, E. *et al. Mol. Syst. Biol.* **5**, 269 (2009).
12. Miao, H. *et al. PLoS ONE* **3**, e2847 (2008).
13. Wang, H. *et al. J. Proteome Res.* **5**, 361–369 (2006).
14. Harris, L.W. *et al. PLoS ONE* **3**, e3964 (2008).