D846–D851 *Nucleic Acids Research, 2009, Vol. 37, Database issue* *Published online 4 November 2008*
doi:10.1093/nar/gkn700

G2Cdb: the Genes to Cognition database

Mike D. R. Croning[1], Michael C. Marshall[1], Peter McLaren[1], J. Douglas Armstrong[2] and Seth G. N. Grant[1,*]

[1]Genes to Cognition Programme, Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA and [2]Institute for Adaptive and Neural Computation, School of Informatics, University of Edinburgh, 10 Crichton Street, Edinburgh EH8 9AB, UK

Received August 15, 2008; Revised September 25, 2008; Accepted September 26, 2008

ABSTRACT

Neuroscience databases linking genes, proteins, (patho)physiology, anatomy and behaviour across species will be valuable in a broad range of studies of the nervous system. G2Cdb is such a neuroscience database aiming to present a global view of the role of synapse proteins in physiology and disease. G2Cdb warehouses sets of genes and proteins experimentally elucidated by proteomic mass spectroscopy of signalling complexes and proteins biochemically isolated from mammalian synapse preparations, giving an experimentally validated definition of the constituents of the mammalian synapse. Using automated text-mining and expert (human) curation we have systematically extracted information from published neurobiological studies in the fields of synaptic signalling electrophysiology and behaviour in knockout and other transgenic mice. We have also surveyed the human genetics literature for associations to disease caused by mutations in synaptic genes. The synapse proteome datasets that G2Cdb provides offer a basis for future work in synapse biology and provide useful information on brain diseases. They have been integrated in a such way that investigators can rapidly query whether a gene or protein is found in brain-signalling complex(es), has a phenotype in rodent models or whether mutations are associated with a human disease. G2Cdb can be freely accessed at http://www.genes2cognition.org.

INTRODUCTION

Synapses are the fundamental unit of computation in the brain playing key roles in information processing, behaviour and disease. They not only transmit information between cells but also detect patterns of neural activity and process this information by activating intracellular biochemical signalling pathways, which subsequently changes the properties of the neuron. G2Cdb presents an integrated view of the role of synapses, focusing on large, high-quality datasets describing synaptic proteins and diseases of the nervous system, particularly those affecting cognition.

Since the year 2000, proteomic studies have increased the number of known synaptic proteins over 10-fold and provided lists of proteins that represent the draft synapse proteome (1,2 and other references therein). The synapse has different compartments, such as the post-synaptic proteome (PSP), comprising ~1100 proteins, and pre-synaptic vesicles with ~80 proteins (3). The high degree of complexity was unexpected and understanding the function of individual proteins and the overall organization of the molecular networks presents a major challenge.

G2Cdb aims to be the central database for warehousing data on the synaptic proteome. Other useful databases exist for molecular neuroscience of which the closest is Synapse DataBase (SynDB). G2Cdb differs fundamentally from SynDB both in terms of the data content and in the way it is constructed. SynDB employs keyword and ontology-term searching of protein sequence and motif databases to provide an informatic definition of the synapse (4). In contrast, G2Cdb uses data curated from published studies of synaptic protein profiling to provide an experimentally validated representation of the mammalian synapse. Both approaches, and thus databases, are highly complementary.

Building upon this proteomic definition of the synapse G2Cdb integrates mouse and human genomic annotation resources, forming the basis of a 'molecular catalogue' of mammalian synaptic genes. Information mined from the human genetics literature reporting associations between synaptic gene mutations and disease is included, as is our in-depth and on-going survey of the neurobiological phenotypes observed in published studies of knockout and other transgenic mice.

With the aim of presenting a global view of the role of synapses in physiology and disease, these datasets have been integrated in a gene-centric manner. The resulting database, G2Cdb, should be of interest to all

*To whom correspondence should be addressed. Tel: +44 1223 834244; Fax: +44 1223 494919; Email: sg3@sanger.ac.uk

neuroscientists, clinicians and geneticists with interests in disease, given the ever-increasing number of synaptic proteins that are involved in human brain diseases such as Alzheimer's disease, autism, mental retardation and schizophrenia (5).

G2Cdb can be freely accessed at www.genes2cognition.org.

CONSTRUCTION AND CONTENTS

G2Cdb consists of a catalogue of experimentally validated mammalian synapse genes that we have integrated with genomic annotation resources, studies of naturally occurring gene mutations found in the human central nervous system (CNS) diseases, brain gene expression resources, as well as behavioural studies and electrophysiological studies using genetically modified mice. The database creation was split into six principle steps (i) compiling a catalogue of mammalian synapse genes from proteomic studies, (ii) attaching gene symbols, protein names and their synonyms to the genes, (iii) constructing gene lists to represent the constituents of synaptic protein complexes and organelles, (iv) linking to external genome and molecular resources, (v) linking to transcript and protein expression resources, and (vi) automated text-mining and then human (manual) curation of published literature. Details for each of these are presented in the following sections.

Compiling a catalogue of mammalian synapse genes

The proteomic study of synaptic organelles and receptor–protein complexes biochemically isolated from nervous tissue gives us the best current picture of the components of the mammalian synapse. These studies utilize a series of purification strategies, followed by liquid chromatography tandem mass spectroscopy and/or large scale immunoblotting to identify proteins (6).

We have extracted the data from seven large-scale proteomic profiling studies to populate G2Cdb (Table 1), thus creating a comprehensive set of proteins found at the mammalian synapse. In total, more than 1100 proteins have been identified revealing an unexpected complexity to the synapse which is also reflected in its complex evolutionary origin (7). We refer to this set as the PSP (2) and this is the current focus of G2Cdb.

As the next step in creating a mammalian synapse gene catalogue, we took the set of 1124 proteins found in the PSP and mapped the protein sequences back to the mouse genome via the Ensembl genome annotation database. Next, we used all-versus-all protein and transcript FASTA (8) searches to ensure the resulting mouse gene set was non-redundant and then assigned unique identifiers to these genes (of the form 'Gxxxxxxxx' where x is a digit 0–9). We found unique identifiers were necessary as genome database identifiers changing regularly in mammalian genomes, particularly in difficult-to-predict, multiexon genes. Having our own identifiers insulates G2Cdb from such changes.

We also store and display homologous gene information for all G2Cdb genes. To date, we have focused on orthology information, utilizing the automated Ensembl predictions as a starting point. We found it valuable to manually curate gene orthology information particularly where clear 1:1 relationships between the human and mouse genes could not be automatically obtained. Such orthologues were manually assigned by searching using the mouse and human gene names on the Mouse Genome Informatics (MGI; 9) and HUGO Gene Nomenclature Committee (HGNC; 10) database websites, respectively. Where this did not provide a match, sequence similarity was used to assign orthologues.

Gene symbols, protein names and their synonyms

Different research communities (e.g. neuroscience or genetics) for historical reasons have often favoured the use of different names or symbols when referring to the same biological entities. We address this by labelling genes with approved gene symbols and additionally stored terms and synonyms obtained from various sources including the gene nomenclature committees (MGI and HGNC) and standardized protein names (UniProt; 11). This means the user can select from either standard names or use common synonyms employed in their field when searching.

Table 1. Published synapse proteomic profiling datasets used in the production of G2Cdb

Paper	Year	Reference
Molecular characterization and comparison of the components and multiprotein complexes in the post-synaptic proteome.	2006	Collins *et al.* (2)
Identification and verification of novel rodent post-synaptic density proteins.	2004	Jordan *et al.* (30)
Proteomics analysis of rat brain post-synaptic density. Implications of the diverse protein functional groups for the integration of synaptic physiology.	2004	Li *et al.* (31)
Semiquantitative proteomic analysis of rat forebrain post-synaptic density fractions by mass spectrometry.	2004	Peng *et al.* (32)
Molecular constituents of the post-synaptic density fraction revealed by proteomic analysis using multidimensional liquid chromatography-tandem mass spectrometry.	2004	Yoshimura *et al.* (33)
Identification of activity-regulated proteins in the post-synaptic density fraction.	2002	Satoh *et al.* (34)
Identification of proteins in the post-synaptic density fraction by mass spectrometry.	2000	Walikonis *et al.* (35)

Representation of synaptic protein complexes and organelles

The protein products of nervous system genes are far from evenly distributed at nerve terminals or synapses. Rather, they are integrated into large macromolecular protein complexes by scaffolding proteins, producing units that have distinct functional roles which create the ability of molecules to couple via adaptors activating various cell-signalling pathways (12). To capture this structural and organizational information in G2Cdb, genes whose protein products are associated may be grouped into one or more gene lists. Prime examples include the two glutamate receptor complexes, the NMDA receptor complex (NRC/MASC) and the AMPA receptor complex (ARC), both of which are key receptor complexes in synaptic transmission and synaptic plasticity at the post-synaptic membrane (1–2). A further example is the clathrin-coated vesicle containing proteins mediating receptor endocytosis (13).

Towards integrating information on the function, interactions and phylogeny of individual proteins within the NRC/MASC complex, we previously probed its organization and developed a model of its function using a combination of annotation, network and statistical approaches (14). We proposed a modular network with 13 distinct subcomponents with distributed functionality that could explain many of the features of synaptic signalling. To facilitate use of our prior analysis and to allow for expansion, we added the gene lists in G2Cdb for the 13 subcomponents of the NRC/MASC complex.

External genome and molecular resources

Links were established for each G2Cdb gene to publicly available genome and molecular resources. These include Ensembl (15), Vega (16), UniProt (11), Entrez Gene (17), GeneCards (18) and OMIM (19) and PubMed (20; see Table 2 for the full list). We used mapping information from Ensembl and/or the approved gene symbols from the various databases as a starting point for the cross-referencing. Cases of ambiguous or absent mapping were manually resolved using the resources' respective websites before acceptance.

Transcript and protein expression resources

Knowledge of the anatomical location and level of a protein expression (distribution) profile may be vital when trying to predict or interpret a phenotype in an animal created by ablating a gene, or while attempting to elucidate the role of a gene product in a particular behaviour.

We have mined the plethora of publicly available gene and protein expression resources for this valuable information. For example, through collaboration with the Allen Brain Institute, we provide the Allen Brain Atlas (21) data links to their database of gene expression patterns in the mouse brain. Other mouse gene expression resources were mined by automated means (BGEM, GenePaint, GENSAT, EMAGE; 22–25). Links to human protein expression profiles are provided for the Human Protein Atlas (26). Using automated cross-referencing these gene and protein expression links are regularly updated, maintaining synchronization with these external resources.

Text-mining and human curation of published literature

Curated databases containing published literature provide an invaluable tool to researchers by organizing and presenting disparate information in a systematic and accessible manner. In the field of synaptic signalling, functional studies have centred on the cellular mechanisms and behavioural roles of synaptic plasticity

In order to optimize the arduous task of performing an extensive literature survey in areas where there are many hundreds of papers published each year, we deployed a bespoke text-mining system (described in refs 5 and 27). Briefly, we pre-processed and indexed a local copy of Pubmed using Lucene (28). We then implemented a basic language classifier using the regular expression capabilities of Perl to extract and rank-order abstracts with likely terms, e.g. 'interacts with' or 'long term potentiation'.

The resulting papers were separated into two lists and then human-curated to confirm both their relevance and the identity of the gene(s) investigated. For one list, we also manually extracted the experimental conditions employed for the synaptic plasticity papers

Table 2. G2Cdb cross-referenced gene, genome, expression and literature resources

Content	Database	URL
Mouse and human genomic annotation	Ensembl	http://www.ensembl.org
Mouse and human genomic annotation	NCBI Entrez Gene	http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene
Human genes and genetic phenotypes	OMIM	http://www.ncbi.nlm.nih.gov/omim
Human-curated gene structures	Vega	http://vega.sanger.ac.uk
Human gene information	GeneCards	http://www.genecards.org
Mouse brain gene expression	Allen Brain Atlas	http://www.brain-map.org
Mouse nervous system gene expression	BGEM	http://www.stjudebgem.org
Mouse gene expression in development	EMAGE	http://genex.hgu.mrc.ac.uk/Emage/database
Mouse embryo gene expression	GenePaint	http://www.genepaint.org
Mouse CNS gene expression	GENSAT	http://www.ncbi.nlm.nih.gov/projects/gensat
Human protein distribution	Human Protein Atlas	http://www.proteinatlas.org
Protein knowledgebase	UniProt	http://www.uniprot.org
Human gene nomenclature	HGNC	http://www.genenames.org
Mouse gene nomenclature	MGI	http://www.informatics.jax.org
Published scientific literature	PubMed	http://www.ncbi.nlm.nih.gov/pubmed

(temperature, stimulating protocol, etc.). For the second list, the behavioural assessment of animals with a gene mutation was manually curated and classed, based on a controlled vocabulary and an anatomically based grouping of the behavioural tests. Presently, we have amassed information on a total of ~340 genes, many of which show phenotypes, by extracting information from about 600 mouse synaptic plasticity and behavioural studies.

Using similar techniques we have also surveyed the human genetics literature looking for associations to disease caused by mutations in synapse genes in our catalogue. To date, we have curated information on 324 diseases, including 90 nervous system diseases, associated with mutations in about 90 synapse genes, from almost 800 articles. Mutations are classified as one of about 35 types, such as single nucleotide polymorphism (SNP), insertion, deletion, nonsense and copy number variation (CNV).

The Genes to Cognition research programme (29) is currently performing a large-scale study of mice carrying mutations in post-synaptic proteins, with extensive phenotyping in biochemical, electrophysiological and behavioural domains. This data is deposited as it becomes available in G2Cdb.

WEB INTERFACE

Public access to G2Cdb is provided through a web interface where users can interactively query and retrieve data. We have integrated the above datasets from disparate knowledge domains so that interrogation by a single text search is possible. Users can rapidly determine if a particular gene is found in brain-signalling complex(es), has been altered and studied in experimental paradigms of learning and memory, and whether a mutation in this gene can be associated with a human disease. As an example, we can take 'SAP102'. Searching for this gene returns links to the G2Cdb 'GeneView' pages for both the mouse and human genes with their species-specific links to external resources. Also (at the time of writing) curated results from two human genetics studies that have linked mutations in the gene encoding SAP102 (*DLG3*) to X-linked mental retardation are retrieved, as is data extracted from a published knockout mouse study reporting electrophysiological characterization (synaptic transmission, short- and long-term plasticity) and results from a battery of behavioural tests (Figure 1).

The website provides an interactive comparison tool called 'CompareGeneLists' to fully exploit the growing number of gene lists loaded into G2Cdb. This tool allows users to select two gene lists and discover the

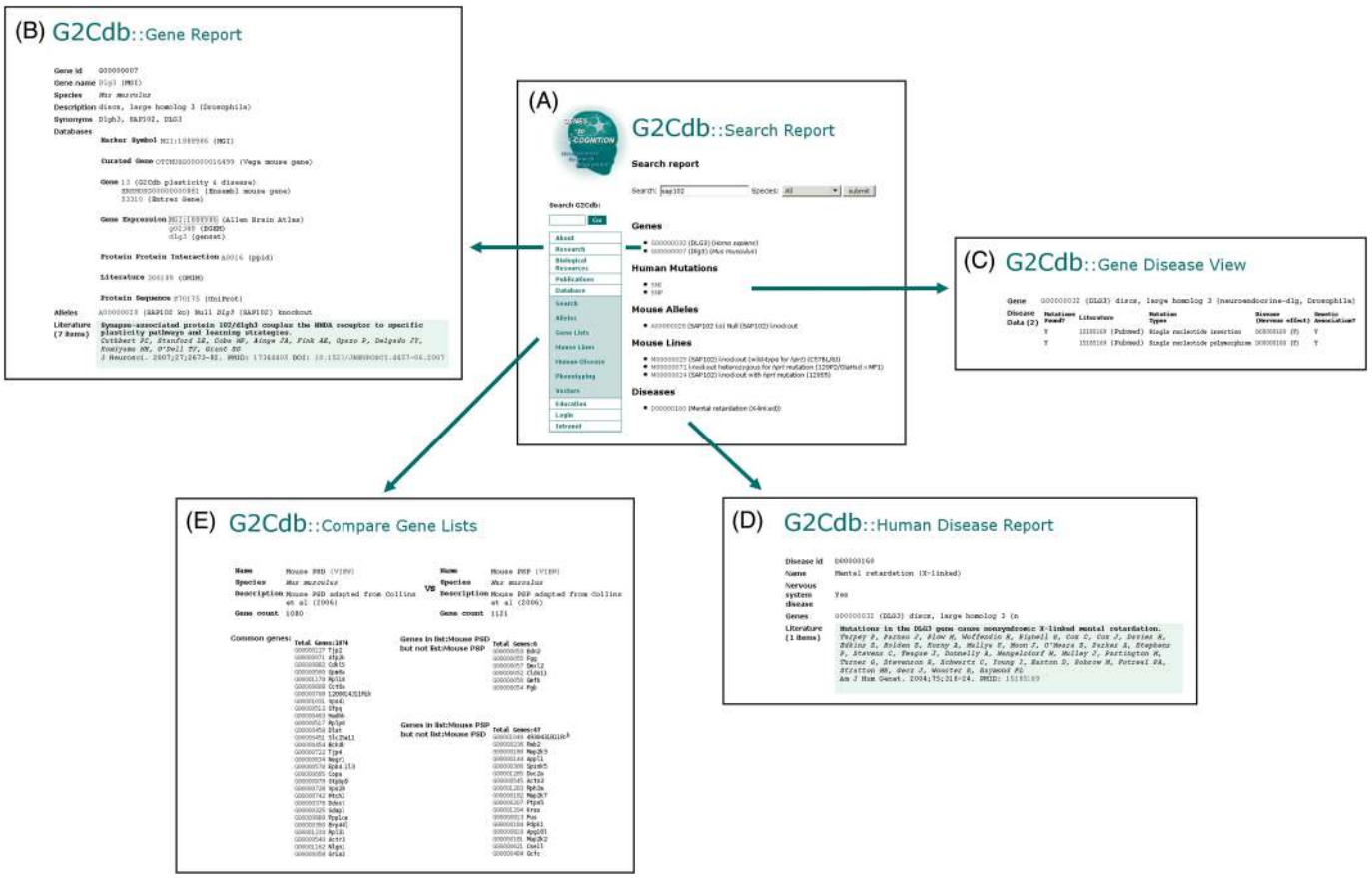

Figure 1. Sample screenshots of G2Cdb. Searching for SAP102 (**A**) returns links to the G2Cdb 'GeneView' pages for both the mouse (**B**) and human genes with their species-specific links to external resources. Also returned are curated results from two human genetics studies (**C**) that have linked mutations in the gene encoding SAP102 (DLG3) to X-linked mental retardation (**D**), as is data extracted from a published knockout mouse study. The gene list comparison tool is also shown, displaying the results of comparing the mouse PSD and PSP datasets (**E**).

common genes and those unique in each set. One can thus pose the question, 'Which are the common and differing genes found to encode the proteins of the NRC/MASC and post-synaptic density?'

Complementing the text-searching abilities of G2Cdb, one can browse the contents of the database by clicking the various links available under 'Database' (on the menu present on the left of all the G2Cdb pages) and in the 'Top Searches' links. Examples are 'Gene Lists', 'Human Disease' and 'Genes shown to affect long-term potentiation (LTP)'.

SOFTWARE IMPLEMENTATION

G2Cdb is implemented in a relational database management system (MySQL) in a schema with about 70 tables in which all the biological information and related controlled vocabularies (CVs) are stored. The database back-end is accessed by an object-orientated Perl API, together with an extensive set of Perl scripts for loading the database. These automated loading scripts integrate our curated literature datasets and those obtained from external resources and perform quality control (QC) checks. The G2Cdb website is driven by a set of complementary CGI scripts.

The principal objects stored by G2Cdb are genes, alleles, genetically modified mouse lines, phenotyping experiments and diseases. The API allows one to create, link, store and retrieve these objects in a high-level manner (independent of the underlying SQL) making it straightforward to implement quite sophisticated analyses. Generic mechanisms exist to attach synonyms, textual notes, literature citations, external databases' cross-references and binary files (such as media files or pdfs) to the core database objects, and to validate object attributes with CVs. The website also supports secure login (by single sign-on) should this be required.

The software package comprising G2Cdb will be useful to other investigators working in research fields outside of cognition who have the need to integrate similar types of data from genomics and experimental genetics (genetically altered mice and phenotyping experiments) with disease and mutation information from human genetics studies. In common with other Wellcome Trust-funded informatics initiatives, the software is open source, and is available free of charge under the terms of the Perl Artistic License (by request to webmaster@genes2cognition.org). The majority of programmes and modules are documented with embedded Perl documentation (POD) and the schema and API are designed very much in keeping with those of Ensembl, so bioinformaticians familiar with this project should find it straightforward to employ the system.

DATA UPDATES

We currently update G2Cdb 2–4 times a year. At these times we release newly incorporated data, make improvements to the website to support retrieval of the new data types and generally improve site usability. We plan to load new synapse proteomic profiling datasets as they become available and continue our text mining and literature curation, focussing on extending the set of genes for which we have curated synaptic plasticity and behavioural studies in genetically modified mouse models. Similarly, we will continue to survey the human genetics literature for mutations in the human orthologues of these genes and their association to disease. The G2C research programme and other large-scale research programmes will provide major volumes of data to G2Cdb. The longer term objective is to provide comprehensive information on all synapse proteins across a wide range of phenotypes, functions and species.

DISCUSSION

Driven by the experimental progress in synapse proteomics and functional studies of these proteins, we identified a need for a specialist database for the organization and function of the synapse proteome. This includes capturing detailed information from a number of sources of experimentally validated results, expert curated information from the literature and links to related external databases. Importantly, G2Cdb provides a mechanism for the effective re-use and analysis of a range of datasets that are expensive to curate/produce by the wider community. The synapse proteome datasets that G2Cdb provides thus offer a basis for future research in synapse biology and provide useful information on brain diseases.

A major application of G2Cdb will be the assembly of molecular networks of the synapse proteome. These include transcriptional, protein interaction and phosphorylation networks. Combining this with phenotypic data on specific proteins in electrophysiology and behaviour from mice, and human disease information will be useful for the systems biology of the synapse.

Toward the public understanding of science and the education of school and undergraduates, we are collaborating with the Dolan DNA Learning Centre at Cold Spring Harbor Laboratory to provide G2Cdb in a format of use to school and college students. An educational website (www.g2conline.org) will cover a broad range of educational material on the subjects of genes and behaviour including novel network representations of the G2Cdb datasets.

ACKNOWLEDGEMENTS

The authors would like to thank Phil Elliot and Keri-Lee Page for their contributions of data to G2Cdb and Lianne Stanford for critical review of the article.

FUNDING

Genes to Cognition Programme (Wellcome Trust). Funding for open access charges: Wellcome Trust.

Conflict of interest statement. None declared.

REFERENCES

1. Husi,H., Ward,M.A., Choudhary,J.S., Blackstock,W.P. and Grant,S.G. (2000) Proteomic analysis of NMDA receptor-adhesion protein signaling complexes. *Nat. Neurosci.*, **3**, 661–669.
2. Collins,M.O., Husi,H., Yu,L., Brandon,J.M., Anderson,C.N., Blackstock,W.P., Choudhary,J.S. and Grant,S.G. (2006) Molecular characterization and comparison of the components and multiprotein complexes in the postsynaptic proteome. *J. Neurochem.*, **97 (Suppl 1)**, 16–23.
3. Takamori,S., Holt,M., Stenius,K., Lemke,E.A., Gronborg,M., Riedel,D., Urlaub,H., Schenck,S., Brugger,B., Ringler,P. *et al.* (2006) Molecular anatomy of a trafficking organelle. *Cell*, **127**, 831–846.
4. Zhang,W., Zhang,Y., Zheng,H., Zhang,C., Xiong,W., Olyarchuk,J.G., Walker,M., Xu,W., Zhao,M., Zhao,S. *et al.* (2007) SynDB: a Synapse protein DataBase based on synapse ontology. *Nucleic Acids Res.*, **35**, D737–D741.
5. Grant,S.G., Marshall,M.C., Page,K.L., Cumiskey,M.A. and Armstrong,J.D. (2005) Synapse proteomics of multiprotein complexes: en route from genes to nervous system diseases. *Hum. Mol. Genet.*, **14 (Spec No. 2)**, R225–R234.
6. Grant,S.G. (2006) The synapse proteome and phosphoproteome: a new paradigm for synapse biology. *Biochem. Soc. Trans.*, **34**, 59–63.
7. Emes,R.D., Pocklington,A.J., Anderson,C.N., Bayes,A., Collins,M.O., Vickers,C.A., Croning,M.D., Malik,B.R., Choudhary,J.S., Armstrong,J.D. *et al.* (2008) Evolutionary expansion and anatomical specialization of synapse proteome complexity. *Nat. Neurosci.*, **11**, 799–806.
8. Pearson,W.R. (1990) Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods Enzymol.*, **183**, 63–98.
9. Bult,C.J., Eppig,J.T., Kadin,J.A., Richardson,J.E., Blake,J.A. and Mouse Genome Database Group, (2008) The Mouse Genome Database (MGD): mouse biology and model systems. *Nucleic Acids Res.*, **36**, D724–D728.
10. Bruford,E.A., Lush,M.J., Wright,M.W., Sneddon,T.P., Povey,S. and Birney,E. (2008) The HGNC Database in 2008: a resource for the human genome. *Nucleic Acids Res.*, **36**, D445–D448.
11. The UniProt Consortium (2008) The universal protein resource (UniProt). *Nucleic Acids Res.*, **36**, D190–D195.
12. Sheng,M. and Hoogenraad,C.C. (2007) The postsynaptic architecture of excitatory synapses: a more quantitative view. *Annu. Rev. Biochem.*, **76**, 823–847.
13. Mills,I.G. (2007) The interplay between clathrin-coated vesicles and cell signalling. *Semin. Cell Dev. Biol.*, **18**, 459–470.
14. Pocklington,A.J., Cumiskey,M., Armstrong,J.D. and Grant,S.G. (2006) The proteomes of neurotransmitter receptor complexes form modular networks with distributed functionality underlying plasticity and behaviour. *Mol. Syst. Biol.*, **2**, 2006.0023.
15. Flicek,P., Aken,B.L., Beal,K., Ballester,B., Caccamo,M., Chen,Y., Clarke,L., Coates,G., Cunningham,F., Cutts,T. *et al.* (2008) Ensembl 2008. *Nucleic Acids Res.*, **36**, D707–D714.
16. Wilming,L.G., Gilbert,J.G., Howe,K., Trevanion,S., Hubbard,T. and Harrow,J.L. (2008) The vertebrate genome annotation (Vega) database. *Nucleic Acids Res.*, **36**, D753–D760.
17. Maglott,D., Ostell,J., Pruitt,K.D. and Tatusova,T. (2007) Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.*, **35**, D26–D31.
18. Safran,M., Chalifa-Caspi,V., Shmueli,O., Olender,T., Lapidot,M., Rosen,N., Shmoish,M., Peter,Y., Glusman,G., Feldmesser,E. *et al.* (2003) Human Gene-Centric Databases at the Weizmann Institute of Science: GeneCards, UDB, CroW 21 and HORDE. *Nucleic Acids Res.*, **31**, 142–146.
19. McKusick,V.A. (2007) Mendelian inheritance in man and its online version, OMIM. *Am. J. Hum. Genet.*, **80**, 588–604.
20. Toouli,J. (2008) PubMed Central. *HPB*, **10**, 3.
21. Lein,E.S., Hawrylycz,M.J., Ao,N., Ayres,M., Bensinger,A., Bernard,A., Boe,A.F., Boguski,M.S., Brockway,K.S., Byrnes,E.J. *et al.* (2007) Genome-wide atlas of gene expression in the adult mouse brain. *Nature*, **445**, 168–176.
22. Magdaleno,S., Jensen,P., Brumwell,C.L., Seal,A., Lehman,K., Asbury,A., Cheung,T., Cornelius,T., Batten,D.M., Eden,C. *et al.* (2006) BGEM: an in situ hybridization database of gene expression in the embryonic and adult mouse nervous system. *PLoS Biol.*, **4**, e86.
23. Visel,A., Thaller,C. and Eichele,G. (2004) GenePaint.org: an atlas of gene expression patterns in the mouse embryo. *Nucleic Acids Res.*, **32**, D552–D556.
24. Heintz,N. (2004) Gene expression nervous system atlas (GENSAT). *Nat. Neurosci.*, **7**, 483.
25. Venkataraman,S., Stevenson,P., Yang,Y., Richardson,L., Burton,N., Perry,T.P., Smith,P., Baldock,R.A., Davidson,D.R. and Christiansen,J.H. (2008) EMAGE—Edinburgh Mouse Atlas of Gene Expression: 2008 update. *Nucleic Acids Res.*, **36**, D860–D865.
26. Berglund,L., Bjorling,E., Oksvold,P., Fagerberg,L., Asplund,A., Al-Khalili Szigyarto,C., Persson,A., Ottosson,J., Wernerus,H., Nilsson,P. *et al.* (2008) A gene-centric human protein atlas for expression profiles based on antibodies. *Mol. Cell Proteomics*, **7**, 2019–2027.
27. Armstrong,J.D., Pocklington,A.J., Cumiskey,M.A. and Grant,S.G. (2006) Reconstructing protein complexes: from proteomics to systems biology. *Proteomics*, **6**, 4724–4731.
28. Hatcher,E. and Gospodnetic,O. (2004) *Lucene In Action*, Manning Publishers, Greenwich.
29. Grant,S.G.N. (2003) An integrative neuroscience programme linking mouse genes to cognition and disease. In Plomin,R., Defries,J.C., Craig,I.W. and McGuffin,P. (eds), *Behavioural Genetics in the Post Genomic Era*, APA Books, Washington, DC, pp. 123–138.
30. Jordan,B.A., Fernholz,B.D., Boussac,M., Xu,C., Grigorean,G., Ziff,E.B. and Neubert,T.A. (2004) Identification and verification of novel rodent postsynaptic density proteins. *Mol. Cell Proteomics*, **3**, 857–871.
31. Li,K.W., Hornshaw,M.P., Van Der Schors,R.C., Watson,R., Tate,S., Casetta,B., Jimenez,C.R., Gouwenberg,Y., Gundelfinger,E.D., Smalla,K.H. *et al.* (2004) Proteomics analysis of rat brain postsynaptic density. Implications of the diverse protein functional groups for the integration of synaptic physiology. *J. Biol. Chem.*, **279**, 987–1002.
32. Peng,J., Kim,M.J., Cheng,D., Duong,D.M., Gygi,S.P. and Sheng,M. (2004) Semiquantitative proteomic analysis of rat forebrain postsynaptic density fractions by mass spectrometry. *J. Biol. Chem.*, **279**, 21003–21011.
33. Yoshimura,Y., Yamauchi,Y., Shinkawa,T., Taoka,M., Donai,H., Takahashi,N., Isobe,T. and Yamauchi,T. (2004) Molecular constituents of the postsynaptic density fraction revealed by proteomic analysis using multidimensional liquid chromatography-tandem mass spectrometry. *J. Neurochem.*, **88**, 759–768.
34. Satoh,K., Takeuchi,M., Oda,Y., Deguchi-Tawarada,M., Sakamoto,Y., Matsubara,K., Nagasu,T. and Takai,Y. (2002) Identification of activity-regulated proteins in the postsynaptic density fraction. *Genes Cells*, **7**, 187–197.
35. Walikonis,R.S., Jensen,O.N., Mann,M., Provance,D.W. Jr., Mercer,J.A. and Kennedy,M.B. (2000) Identification of proteins in the postsynaptic density fraction by mass spectrometry. *J. Neurosci.*, **20**, 4069–4080.