nature neuroscience

Evolutionary expansion and anatomical specialization of synapse proteome complexity

Richard D Emes[1,6], Andrew J Pocklington[2,6], Christopher N G Anderson[3,6], Alex Bayes[3], Mark O Collins[3], Catherine A Vickers[4,5], Mike D R Croning[3], Bilal R Malik[2], Jyoti S Choudhary[3], J Douglas Armstrong[2] & Seth G N Grant[3]

Understanding the origins and evolution of synapses may provide insight into species diversity and the organization of the brain. Using comparative proteomics and genomics, we examined the evolution of the postsynaptic density (PSD) and membrane-associated guanylate kinase (MAGUK)-associated signaling complexes (MASCs) that underlie learning and memory. PSD and MASC orthologs found in yeast carry out basic cellular functions to regulate protein synthesis and structural plasticity. We observed marked changes in signaling complexity at the yeast-metazoan and invertebrate-vertebrate boundaries, with an expansion of key synaptic components, notably receptors, adhesion/cytoskeletal proteins and scaffold proteins. A proteomic comparison of *Drosophila* and mouse MASCs revealed species-specific adaptation with greater signaling complexity in mouse. Although synaptic components were conserved amongst diverse vertebrate species, mapping mRNA and protein expression in the mouse brain showed that vertebrate-specific components preferentially contributed to differences between brain regions. We propose that the evolution of synapse complexity around a core proto-synapse has contributed to invertebrate-vertebrate differences and to brain specialization.

Learning and adaptation to changing environments are properties shared by all animals and may be involved in adaptive radiation of species into environmental niches. Studies in a wide range of multicellular organisms have shown that simple forms of learning found in invertebrates, such as sensitization and habituation, are building blocks for more complex forms found in vertebrates[1,2]. Attempts to uncover the neurobiological basis of behavioral complexity have focused on the differences between vertebrates, where a high degree of behavioral flexibility has evolved (apparently independently) multiple times (for example, in corvids, cetaceans and primates). Of the factors proposed to explain species differences, it has been argued that those reflecting information-processing capacity (number of cortical neurons and conduction velocity of cortical fibers) correlate best with intelligence among mammals[3]. Although synapses are fundamentally involved in neural information processing, discussions of brain and behavioral evolution typically do not consider the possibility of synaptic molecular evolution, of which little is known.

The biology of learning and other cognitive functions involves the activation of neurotransmitter receptors on the postsynaptic side of the synapse by patterns of neuronal activity, triggering biochemical pathways that lead to changes in neuronal function[1]. The PSD contains multiprotein signal-transduction complexes made up of neurotransmitter receptors and associated proteins that are essential for induction of synaptic plasticity and learning[4–7]. Proteomic studies in mice have shown that ionotropic NMDA and metabotropic subtypes of glutamate receptors are linked by scaffold proteins (MAGUKs) into complexes of 186 proteins that are referred to as NMDA receptor complex (NRC) or MASC[8–10]. The PSD itself has a large degree of complexity, involving ~1,000 identified proteins from a wide variety of functional classes[11–13]. A high percentage of MASC and PSD genes possess physiological, behavioral and disease phenotypes, with single gene studies indicating that over 40 MASC proteins are involved in synaptic and behavioral plasticity in rodents and in human brain disorders[14–16] (http://www.genes2cognition.org/db.html). These proteins and complexes provide a suitable template for examining the molecular evolution of learning and synapse organization.

Here we examine synapse evolution using genomic, proteomic and expression profiling of postsynaptic proteins. Genomic comparison of 19 species indicates that there are differences in the complexity and organization of synapses. This was confirmed by proteomic studies of *Drosophila* MASC, which revealed substantial differences in complexity compared with mouse MASC. The expression profiles of postsynaptic proteins in mouse brain showed a relationship between the evolution of synapse proteins and their pattern of expression. We have created a model for the molecular origins and evolutionary diversification of the synapse, highlighting the involvement of molecular complexity in signal processing and behavior.

[1]Institute for Science and Technology in Medicine, Keele University, Thornburrow Drive, Hartshill, Stoke-on-Trent ST4 7QB, UK. [2]Institute for Adaptive and Neural Computation, School of Informatics, University of Edinburgh, 5 Forrest Hill, Edinburgh EH1 2QL, UK. [3]Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK. [4]Division of Neuroscience, University of Edinburgh, 1 George Square, Edinburgh EH8 9JZ, UK. [5]Okinawa Institute of Science and Technology, Neurobiology Research Unit, Suzaki, Uruma, Okinawa, Japan. [6]These authors contributed equally to this work. Correspondence should be addressed to S.G.N.G. (sg3@sanger.ac.uk).

Received 21 November 2007; accepted 7 May 2008; published online 8 June 2008; doi:10.1038/nn.2135

RESULTS

Evolution of the synapse proteome

To investigate the origins of the mammalian synapse proteome, we identified orthologs of 651 genes corresponding to mouse postsynaptic proteins (570 genes from the PSD, 183 from the MASC and 102 that were common to both; see Methods and refs. 8,10) in 19 different species (**Supplementary Tables 1** and **2** online). The species studied comprised a wide range of animals with nervous systems of differing anatomical complexity: invertebrates, nonmammalian vertebrates and mammals. We also identified orthologs in an out-group that does not possess a nervous system, the unicellular eukaryote *Saccharomyces cerevisiae*. Validation was performed (repeating ortholog searches by hand using key pairs of organisms and a deeper substitution matrix) to ensure that the absence of orthologs did not simply reflect sequence divergence between species (**Supplementary Methods** online).

The numbers of PSD and MASC orthologs in each species correlate well with each other ($R^2 = 0.99$), and clear differences were seen between yeast, invertebrates and vertebrates (**Fig. 1a**,**b**). Approximately 23% of all mammalian synapse proteins were detected in yeast (21.2% MASC, 25.0% PSD) and ~45% were detected in invertebrates (46.2% MASC, 44.8% PSD). Therefore, a substantial proportion of genes encoding MASC and PSD orthologs precede the origins of the nervous system, with apparent stepwise expansions following the divergence of metazoans from eukaryotes and vertebrates from invertebrates. The inclusion of a urochordate (*Ciona intestinalis*), the closest living relatives of vertebrates[17], allowed us to locate this second expansion quite precisely, although analysis of a member of the family *Myxinidae* (hagfish) would be necessary to distinguish between vertebrate- and craniate-specific expansion. We would not expect analysis of additional unicellular eukaryotes to substantially alter the picture of metazoan

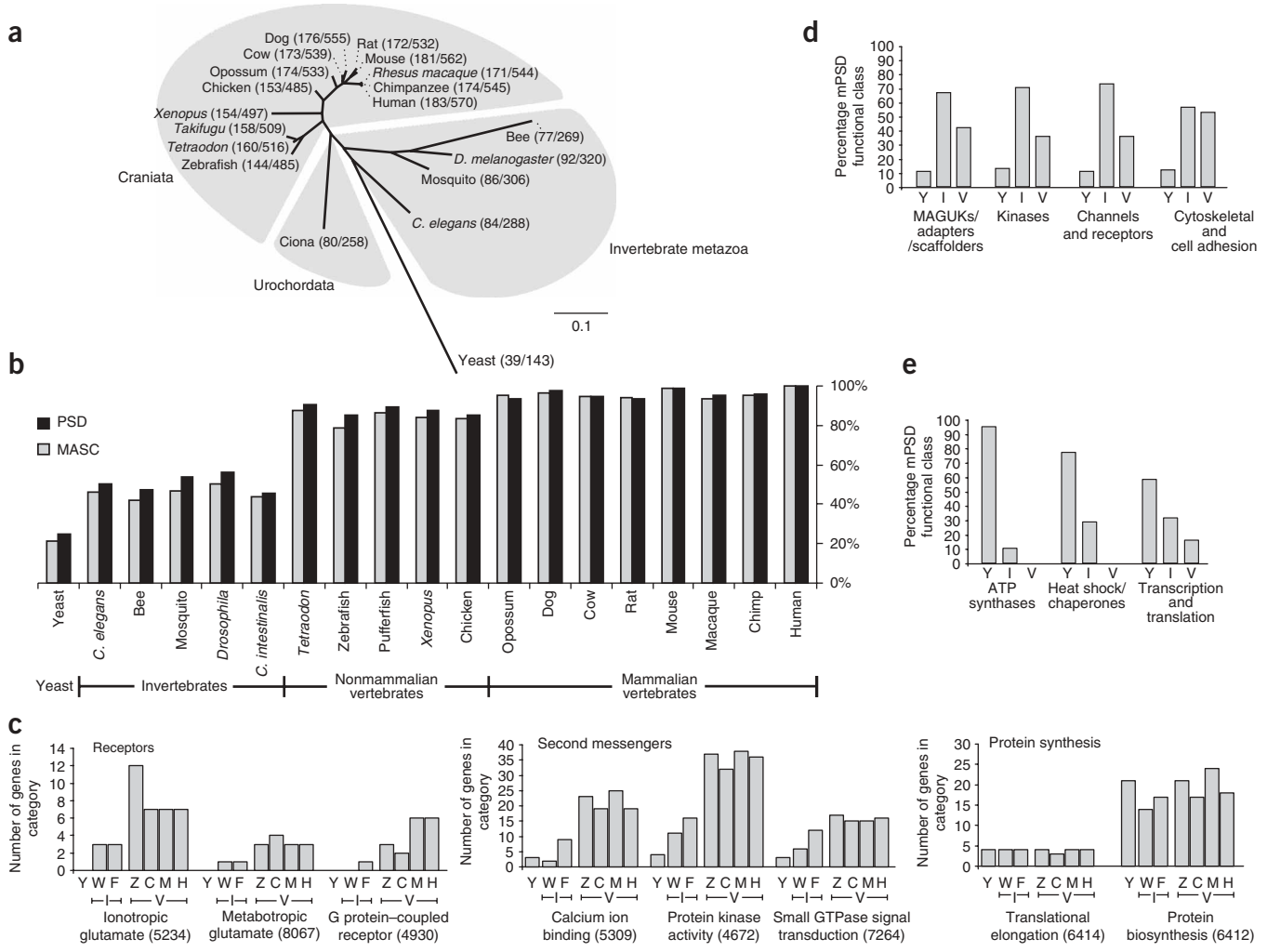

Figure 1 Comparison of PSD and MASC homologs and expansion of selected functional groups of genes. (**a**) The phylogenetic relationships of the species studied are shown. The numbers of NRC/PSD orthologs detected are in parentheses. (**b**) The occurrences of PSD and MASC homologs found in each of the 19 species are shown as a percentage of those found in human. Where annotation of multiple homologs was reported by Ensembl, a single positive hit was recorded. (**c**) Evolution of learning and plasticity mechanisms. From the 651 PSD/MASC genes, we plotted the number of genes in different species (yeast (Y), worm (W), fly (F), zebrafish (Z), chicken (C), mouse (M) and human (H)) that were involved with three major molecular mechanisms (receptors, second messenger signaling, protein synthesis) of learning and memory. For clarity, only data from seven representative species are shown. The numbers in parentheses are the Gene Ontology terms identifiers (http://www.geneontology.org/). Data were obtained from Ensembl. (**d**) Upstream signaling components showed increasing rates of expansion toward mammalian lineage. The proportion of each functional class (as a percentage of the total number of mouse MASC/PSD genes belonging to it) whose earliest identifiable ortholog occurs in yeast, invertebrates (I) or vertebrates (V) is shown. (**e**) Downstream signaling components showed decreasing rates of expansion toward mammalian lineage. The proportion of each functional class (as a percentage of the total number of mouse MASC/PSD genes belonging to it) whose earliest identifiable ortholog occurs in yeast, invertebrates or vertebrates is shown.

expansion, as it primarily involves elaboration of intercellular signaling pathways (fundamental to the existence of metazoa), whereas genes linked to protein synthesis and metabolism are largely identifiable in yeast (see below).

Synaptic plasticity and learning involve the modulation of cellular function and morphology by receptor-associated signaling pathways. To identify unicellular processes from which synapse functionality may have arisen, we considered the function of PSD/MASC orthologs in yeast as described by the *Saccharomyces* Genome Database (http://www.yeastgenome.org/) (**Supplementary Table 3** online). The majority of the orthologs contribute to generic cellular functions (protein synthesis and degradation, vesicular trafficking, and regulation of the actin cytoskeleton) that are regulated in response to environmental factors (ions, nutrients and pheromones). Signal transduction pathways mediating environmental responses contained ~15% of the orthologs. Although (predictably) neurotransmitter receptors were absent, other aspects of synaptic signaling possessed functional counterparts in yeast. Orthologs of calmodulin (*CMD1*) and calcineurin (*CMP2*) control Ca^{2+} homeostasis by regulating transcription of *PMC1* (ref. 16), an ortholog of mouse synaptic Ca^{2+} pump *Atp2b4*. Independent of Ca^{2+}, *CMD1* regulates the actin cytoskeleton (via interaction with the Arp2/3 complex, also found in the PSD) and receptor-mediated endocytosis[18]. Yeast orthologs of mouse *nf1* (*IRA2*), *pka* (*TPK2*), *Erk2* (*FUS3*) and *Gnb5* (*STE4*) belong to major pathways regulating transcription, cell morphology and adhesion downstream of nutrient- and pheromone-sensitive G protein–coupled receptors[19–22]. Therefore, components of synaptic pathways regulating protein synthesis and structural plasticity in rodents have analogous roles in unicellular responses to environmental cues (ions, nutrients) and simple cell-cell (pheromonal) communication.

In parallel with the expansion of the numbers of orthologs, the total number of protein domains detected in each set of orthologs increased (**Supplementary Fig. 1** online). However, the number of domain types did not increase to the same extent, with the difference between invertebrates and vertebrates being much less pronounced (**Supplementary Fig. 1**). These data suggest that synapse proteome expansion does not so much reflect the recruitment of proteins containing new domain types, but more the expansion of protein types already present: that is, innovation by gene family duplication and diversification, rather than by integration or *de novo* generation of new protein types (see **Supplementary Note** and **Supplementary Table 4** online for further analysis of domain types).

We next examined whether specific types of proteins were involved in synaptic proteome expansion. Gene Ontology annotations (http://www.geneontology.org/) were used to evaluate the number of PSD and MASC genes that were associated with synaptic functionality and other more general cell-biological processes in yeast, worm, fly, zebrafish, chicken, mouse and human (**Fig. 1c**). Neurotransmitter receptors were present in all organisms with nervous systems, and vertebrates consistently had greater numbers of receptors than invertebrates. The components of second messenger pathways also showed expansion in vertebrates and invertebrates. In contrast, the representation of protein synthesis machinery did not increase. To ensure that these results do not simply reflect incomplete gene annotation, we took a previously published classification[8] of all 651 synapse proteins into functional families and compared their evolutionary expansion. In agreement with the Gene Ontology term analysis, functional families corresponding to upstream signaling and structural components were poorly represented in yeast, undergoing increasing expansion in invertebrates and vertebrates (**Fig. 1d**). As predicted, the majority of components from downstream cell biological processes could be identified in yeast and early metazoans, and showed substantially reduced expansion (**Fig. 1e**). Of the upstream components, 44% of all cytoskeletal and cell adhesion molecules were vertebrate in origin, significantly more than would be expected from a random sample of genes ($P < 10^{-4}$, see Methods). Only 10% of the cytoskeletal and adhesion genes possessed detectable orthologs in yeast ($P < 10^{-4}$). Also underrepresented in yeast (10% or less identified) were channels and receptors ($P = 0.003$), MAGUKs, adaptors and scaffold proteins ($P = 0.02$) and kinases ($P = 0.05$). Downstream processes comprised the majority of synaptic components with identifiable orthologs in yeast. Notably enriched were genes linked to transcription and translation (over 50%, $P < 10^{-3}$), including ribosomal (85%, $P < 10^{-5}$), protein folding and trafficking (70% of heat shock and chaperones, $P < 10^{-5}$) and metabolism genes (90% of ATP synthases, $P < 10^{-4}$). Therefore, although signal transduction pathways linked to cell surface receptors showed evidence of expansion at the yeast:metazoan and invertebrate:vertebrate boundaries, the downstream cell-biological processes that they regulated did not.

These data suggest that most functional types of synaptic proteins were present in early metazoans and that the proto-synapse constructed from this core functionality has been elaborated on during the evolution of invertebrates and vertebrates. Recent studies confirm that many families of synapse and cell signaling genes are present in the phylum *Porifera* (sponges), supporting the hypothesis that core synaptic signaling components were present at the base of animal kingdom[23–25]. Elaboration appears to have primarily involved gene family expansion and diversification among upstream signaling and structural components (receptors, scaffolding proteins, and cytoskeletal, adhesion and signal transduction molecules).

Key predictions arising from these comparative genomic data are that invertebrate synapses show reduced signaling complexity (number of signaling and structural components) compared with vertebrates, possess components of generic cell-biological processes (for example, protein synthesis and metabolism) that are predominantly of pre-metazoan origin, and show evolutionary expansion in upstream signaling and structural components. To test these predictions, it was necessary to collect proteomic data from invertebrate synapses, and, although PSDs were observed with electron microscopy in the fly brain[26], they have not been analyzed at the molecular level. However, the *discs large* protein (Dlg), which is a *Drosophila* MAGUK and the homolog of mammalian PSD-95, SAP102 and PSD-93, has been studied and is expressed widely in the fly nervous system[27]. Therefore, we isolated *Drosophila* MASC complexes, translating to fly the methods used to isolate mouse MASC.

Isolation of *Drosophila* MAGUK-associated signaling complexes

As per a previously described method for purifying mouse MASCs (mMASC)[28], we generated a C terminus hexapeptide of the *D. melanogaster* NR2 subunit (dPEP6; EMETVL) and a control hexapeptide lacking the PDZ-interaction motif (dPEP6ΔVL; IAEMET, see **Supplementary Methods** and **Fig. 2**) for affinity purification of protein complexes. We found that affinity columns using dPEP6, but not dPEP6ΔVL, bound Dlg (**Fig. 2a**). Moreover, Coomassie-stained gels showed that dPEP6 columns retrieved many proteins and that the Dlg2 band was notably absent from dPEP6ΔVL columns (**Fig. 2b**). We therefore concluded that dPEP6 was capable of isolating fly MAGUK proteins and MASCs (fMASC). We identified 220 fMASC proteins using mass spectrometry (see Methods and **Supplementary Table 5** online), suggesting that *Drosophila* may possess complexes of a size comparable to those found in mouse[8,10]. At first glance, this appeared to contradict the comparative genomic data, but closer inspection

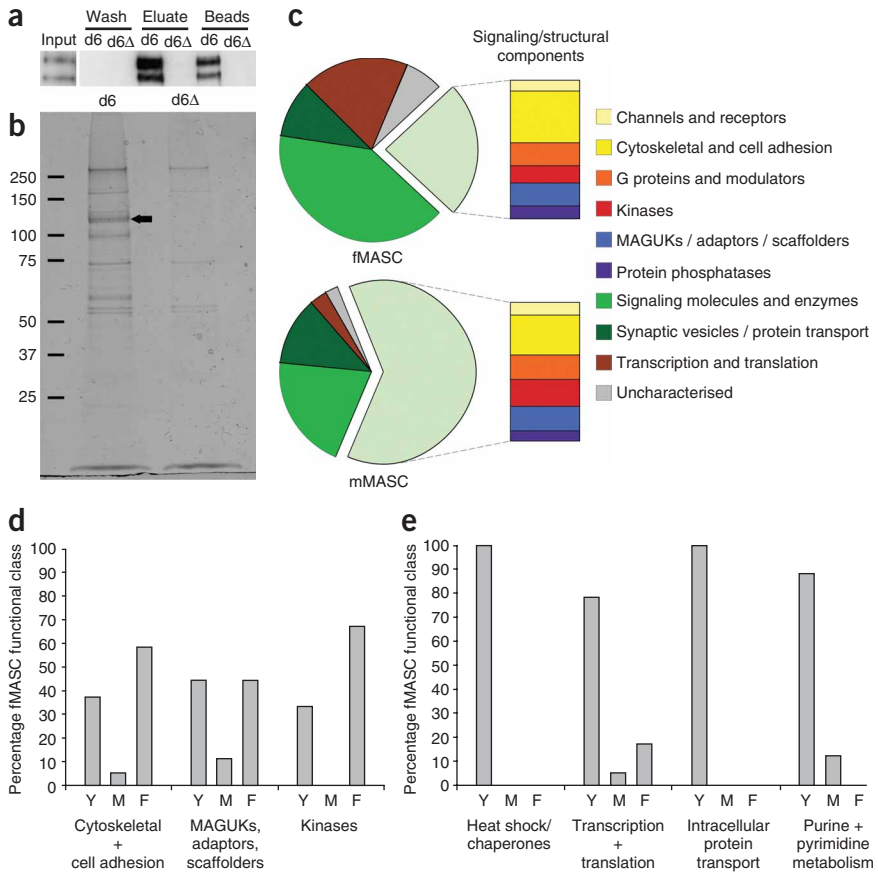

Figure 2 Proteomic analysis of the *Drosophila* MASC. (**a**) Dlg bound the *Drosophila* NR2 C-terminal peptide. Protein extracts (input) of fly heads were incubated with dPEP6 (d6) or dPEP6ΔVL (d6Δ) resin-immobilized peptide. After binding, the resin was washed (wash), bound protein was eluted (eluate) and residual bound protein (beads) was assayed for the presence of *Drosophila* Dlg using immunoblotting. (**b**) Eluates of d6 and d6Δ columns were run on a 4–12% SDS PAGE gel and stained with coomassie brilliant blue. The arrow indicates the region on the gel that gave the majority of Dlg peptides after mass spectrometry. (**c**) Pie charts show the percentage of fMASCs (220 proteins) and mMASCs (186 proteins) belonging to each functional protein class. Key indicates color code for identity of specific classes (downstream, effector components are encapsulated in the green and brown segments). (**d**) Upstream signaling component classes showed significant expansion following divergence of fly and chordate lineages. The proportion of each functional class (as a percentage of the total number of fMASC genes belonging to it) whose earliest identifiable ortholog occurs in yeast, early metazoans (M, common to fly and chordate lineages) or is fly-specific (F) is shown. (**e**) Downstream classes were predominantly of unicellular eukaryotic origin. The proportion of each functional class (as a percentage of the total number of fMASC genes belonging to it) whose earliest identifiable ortholog occurs in yeast, early metazoans or is fly-specific is shown.

revealed major differences in the types of proteins that were present. When fMASC proteins were categorized into functional protein families using the scheme developed for mMASC[8], it was found that upstream signaling and structural components (receptors, scaffolding proteins, signal transduction molecules, etc.) accounted for ~25% of fMASC proteins, compared with >60% of mMASCs (**Fig. 2c**; note that fMASC signaling molecules and enzymes are predominantly metabolic enzymes, heat shock and chaperone proteins, and mitochondrial proteins). Thus the molecular complexity of MASC signaling in fly is roughly half that in mouse (in both relative and absolute terms), as predicted.

To test the remaining predictions, we next asked which types of fMASC components (if any) showed evidence of evolutionary conservation or expansion when compared with yeast. We also noted whether expansion preceded the divergence of fly and mouse lineages (the corresponding genes being detectable in chordates) or whether it was fly specific. Orthologs of fMASC genes were identified in yeast and in the chordate species previously analyzed (**Supplementary Table 6** online). We identified 71% of the fMASC genes in yeast, with only 64 (29%) appearing to be of metazoan origin. Consistent with our predictions, the vast majority of downstream components were present in yeast (**Fig. 2e**), whereas upstream signaling and structural components of fMASCs showed fly-specific expansion (**Fig. 2d**). Thus, both fMASCs and mMASCs appeared to have undergone lineage specific adaptation.

We were interested in the extent to which fMASCs and mMASCs represent similar synaptic subcomponents by comparing the composition of fMASCs to mMASCs and mPSD. We focused on proteins with a primary role in synaptic function and signaling as opposed to protein metabolism and synthesis or other cellular processes (for example, mitochondrial function) (**Supplementary Tables 5** and **7** online). We identified 67 (30%) fMASC proteins as having primarily synaptic function, in contrast with 155 (83%) mMASCs and 355 (62%) mPSDs (85 of which were also present in mMASC; **Supplementary Table 7**). fMASCs and mMASCs appeared to contain similar proportions of each functional family of proteins (suggesting a similar type of signaling complex), whereas fMASCs and mPSDs were less similar. To quantify this, we calculated the probability of a random set of 67 proteins from mMASCs or mPSDs having the same functional representation (number of proteins in each functional class) as those from fMASC. The probability of obtaining the fMASC representation from mMASCs was 80-fold greater than the probability of obtaining it from mPSD, and over 8,000-fold greater than obtaining it from the subset of mPSDs that were not found in mMASC. Thus, in terms of their general composition, fMASCs and mMASCs comprise similar functional subcomponents of the synapse.

Finally, we considered the genetic similarity of fMASCs and mMASCs, and whether they showed evidence of having evolved from a common synaptic subcomponent. Given the extent of both fly-specific and vertebrate-specific adaptation, we would not expect to find a high degree of molecular 1:1 orthology between fMASCs and mMASCs. However, we would expect to find greater evidence of expansion from common ancestral genes amongst synaptic signaling molecules. We identified 44 (20%) fMASC components as being orthologous to 56 mMASC and mPSD genes. Of these, 40 (71%) showed evidence of gene family expansion (one-many and many-many gene mapping) between fly and mouse, including 85% of all fMASC synaptic function gene orthologs ($P = 0.01$). Using BLAST[29] to identify

more distant relationships, we found that 62 (93%) fMASC synaptic function genes had identifiable homologs in the mouse genome, with 50 genes (75%) possessing synaptic function homologs in mMASCs and/or mPSD, (**Supplementary Methods**). Supporting the identification of fMASCs and mMASCs as similar synaptic subcomponents, 33 of the 50 genes were present in mMASC. Thus, in terms of their isolation, functional composition and phylogenetic relationship, fMASCs and mMASCs appear to represent species-specific adaptations of a common synaptic subcomponent, having diverged by duplication, recruitment and replacement of genes.

It seems highly plausible that the marked increase in molecular signaling complexity noted in vertebrates contributes to their increased capacity for behavioral complexity. Even relatively small changes in the number of different signaling components may have a large multiplicative impact on neuronal function. For example, invertebrates have single NR2 and Dlg genes, whereas vertebrates have four of each; the number of potential NR2-Dlg complexes in vertebrates is 16-fold higher. Extending this to other expanded families of synaptic proteins suggests that there is a major increase in the number of combinations with distinct functional properties that are available to the vertebrate nervous system.

Although molecular complexity clearly contributes to neuronal function, it does so in the context of an expanded mammalian nervous system that has undergone substantial regional specialization. Indeed, the genome duplication events leading to more complex synapse proteomes in vertebrates predate the origins of species with anatomically large nervous systems (all vertebrate species had identifiable orthologs of ~80% or more mouse MASC/PSD genes; **Fig. 1**). This immediately leads to the question of whether the complexity of the synapse proteome has been utilized in the expansion and diversification of vertebrate nervous systems, and whether anatomical regionalization has been accompanied by synapse specialization through the tuning of expression of synaptic proteome components.

MASC expression diversity in mouse brain

We examined the expression patterns of synaptic proteins in different regions of the mouse brain. In total, we examined more than 150 molecules and 22 brain regions using protein and mRNA assays. Proteins were examined by western blotting of tissue extracts probed with antibodies to 65 different synaptic proteins, 56 of which are present in the MASC. We carried out immunohistochemistry (IHC) of mouse brain sections with antibodies to 43 different synaptic proteins, 39 of which were MASC proteins. mRNAs were examined by obtaining *in situ* hybridization (ISH) results for 55 MASC genes from the Brain Gene Expression Map (http://www.stjudebgem.org)[30]. Microarray data was obtained for 148 different MASC genes[30]. Western blot, IHC and ISH data was collated for four brain regions (hippocampus, cortex, striatum and cerebellum), whereas microarray data was available for a larger set of 22 regions[31]. To facilitate comparisons between datasets in the four common regions, we used a similar scoring method where each gene in each brain region was assigned an expression score ranging from 0 to 4 (see **Supplementary Methods** and **Note** for correlation values between the different datasets). Expression data are contained in **Supplementary Table 8** online and representative IHC images are presented in **Supplementary Figure 2**.

We found that the vast majority of proteins were clearly co-expressed in each area of the brain that we investigated, with >95% of genes being co-expressed in forebrain structures (hippocampus, striatum and cortex) and >80% being expressed in all four regions. It was apparent by visual inspection of all datasets, however, that there was variation between the expression patterns of these genes (**Supplementary Fig. 2** and **Supplementary Table 8**). We therefore classified expression patterns by anatomical variation of expression (AVex) of proteins and mRNAs using four categories: $AVex^{zero}$ (the same expression score in each region and therefore having zero variation), $AVex^{low}$ (the difference between the highest and lowest expression scores of ≤1, but not zero), $AVex^{high}$ (the difference between the highest and lowest scores ≥3) and $AVex^{med}$ (the expression variability lying between $AVex^{high}$ and $AVex^{low}$). It was clear that the majority of genes and proteins were in the variable categories, and for the microarray, IHC and ISH data, the majority of genes and proteins were in the two most variable AVex classes (**Fig. 3a**). These data indicate that each brain region expressed a similar set of postsynaptic proteins, but in different combinations of expression levels. These profiles, or barcodes, can be used to identify a particular region and are consistent with previous studies[31–33]. These region-specific expression profiles indicate that variation in the expression of synaptic proteome components confers regional specialization with differential signal processing.

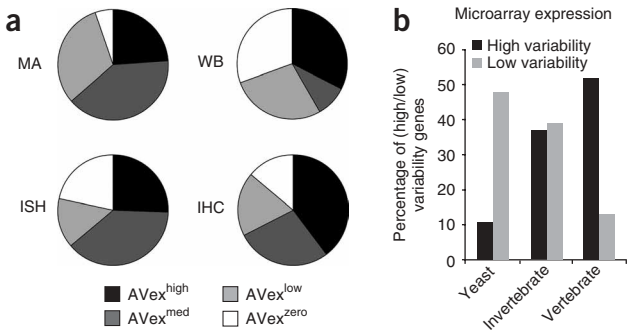

Figure 3 Variation in expression patterns in mouse brain regions. (**a**) Relative proportions of genes with variable expression patterns using four methods. The percentages of genes in each AVex class for each of the methods of testing expression are shown. Note that genes showing no variation in expression between brain regions ($AVex^{zero}$) were in the minority. The expression datasets are (clockwise from top left) microarray (MA), western blot (WB), immunohistochemistry (IHC) and *in situ* hybridization (ISH). Expression variability classes were $AVex^{zero}$ (all expression scores equal), $Avex^{low}$ (scores of only 4 and 3), $AVex^{med}$ (all scores between 4 and >1) and $AVex^{high}$ (scores of 4 and ≤1). (**b**) Variation in brain expression are a function of phylogeny. Shown are the percentages of high (black) and low (gray) variability MASC genes whose earliest identifiable ortholog was present in yeast, invertebrate or vertebrate. Note that the majority of high variability genes were of vertebrate origin and the majority of low variability genes were of premetazoan origin (that is, present in yeast). Microarray data from 22 mouse brain regions was used.

Because proteomic and phylogenetic analyses show that adaptation of synaptic function between species primarily involves the tuning of upstream signal processing pathways, we investigated the relative contribution of pre-metazoan, invertebrate and vertebrate innovations to regional specialization of synaptic function and asked whether specific classes of molecules were involved. Using microarray data covering expression of 148 MASC genes over 22 brain regions, we calculated the standard deviation in each gene's expression and identified those with high (s.d. ≥ 1) or low (s.d. ≤ –1) variability (**Supplementary Methods**). We then plotted the percentage of high and low variability genes found in each of the phylogenetic groupings (**Fig. 3b**). Invertebrate and vertebrate innovations contributed fewer genes to the least variable category, whereas vertebrate innovations contributed greatly to the most variable category. We found that 52% of all highly varying genes were of vertebrate origin, significantly more

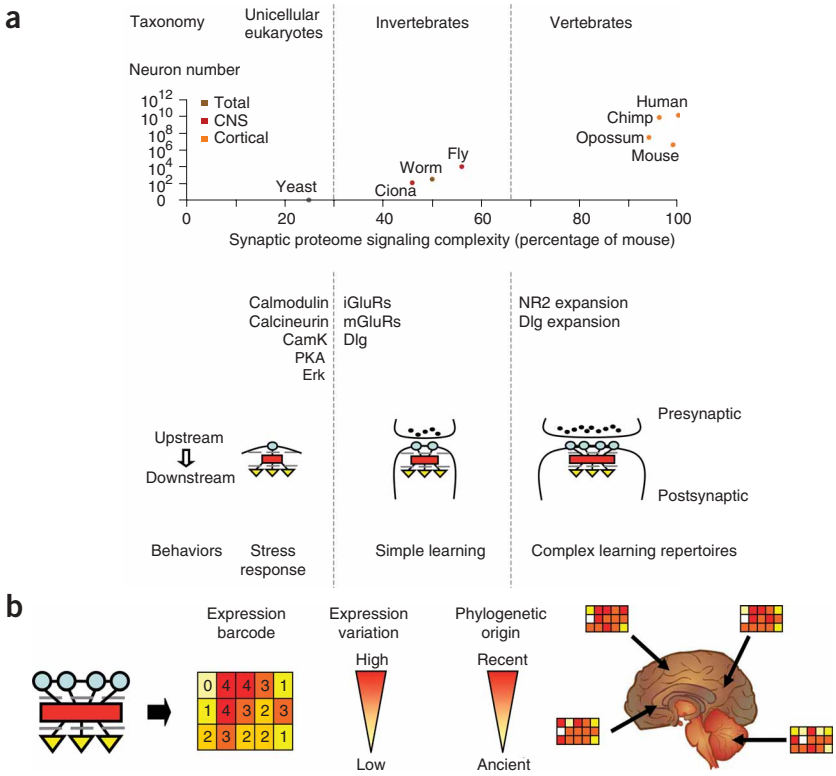

Figure 4 Summary of relationships of synaptic proteome evolution with neuronal number, behavior and expression patterns. (**a**) Relationship of synapse/behavioral complexity to taxonomic grouping of species. In the scatter plot of neuron number against synapse proteome complexity, we obtained estimates of neuron numbers from previous data[3,48–50]. Synaptic proteome complexity was estimated as the percentage of mouse MASC/PSD components possessing orthologs. Several genes of interest to learning and plasticity are listed where they first arise. The schematic representations of signaling complexes use three interlinked shapes (blue circles represent upstream receptor/adhesion proteins, the red box indicates signaling proteins and yellow triangles are downstream proteins). The cell membrane is indicated as a dark line and the pre- and postsynaptic terminal is indicated for invertebrates and vertebrates. The amount of blue circles and the size of the red box increased, illustrating their relative expansion. Behaviors indicated that, although all organisms respond to their environment, the ability to alter these responses and manipulate the environment show marked differences in complexity[2]. Note that the expansion of mammalian brain size occurred after the expansion of synaptic proteome complexity. (**b**) Mammalian MASC complexes and brain region expression variation. A schematic representation of MASC is shown and the expression levels for five proteins from the three levels of MASC are shown (expression barcode). Upstream proteins showed greater variation in expression levels and were of more recent origins. The cartoon of the brain indicates that the expression barcode is distinct for different neuronal populations.

than would be expected from a random sample of genes ($P = 0.003$), and 48% of all least variable genes were of pre-metazoan origin ($P = 0.006$). Similar results were observed in the western blot, IHC and ISH datasets (**Supplementary Fig. 3**).

These data clearly indicate that the genes contributing most to anatomical variation in expression in the mouse brain are typically of more recent origin. It was also found that highly varying genes were substantially enriched with upstream signaling and structural components, reflecting their greater expansion in invertebrates and vertebrates. Of the ten MAGUKs, adaptors and scaffolding proteins for which western blot expression was available, eight were classified as highly varying, accounting for 42% of all $AVex^{high}$ proteins in this dataset ($P = 0.002$), and including all six PDZ-domain containing scaffolding proteins ($P = 0.0008$). Similarly, all seven glutamate receptors studied by IHC were classified as $AVex^{high}$ (41% of all highly varying IHC genes, $P = 0.0007$). In contrast, no members of the signaling molecules and enzymes class (60% of which consisted of ATP synthases, mitochondrial proteins and other enzymes) were found to be highly varying ($P = 0.001$). In the telencephalon, this class of molecules accounted for 42% of all molecules with the least varying microarray expression ($P = 0.006$).

DISCUSSION

Using proteomic, genomic and expression profiling tools, we studied the origin and evolution of brain synapses. The data are consistent with a model in which core components of the synapse originated in unicellular eukaryotes, where they are involved in intracellular signaling and responses to environmental stress (**Fig. 4**). This core may represent the ancestral origins of the brain. Stepwise expansions in molecular signaling complexity coincided with the divergence of metazoans from eukaryotes and vertebrates from invertebrates. Most functional types of synapse protein were present in early metazoans, and elaboration on this core functionality primarily involved gene family expansion and diversification among upstream signaling and structural components (receptors, scaffolding proteins, cytoskeletal, adhesion and signal transduction molecules). Proteomic comparison of vertebrate (mouse) and invertebrate (fly) MASCs showed them to be species-specific adaptations of a common synaptic subcomponent, having diverged by duplication, recruitment and replacement of genes. Thus, substantial expansions in complexity of the synaptic proteome and specialization have occurred during diversification.

Expansion of the synaptic proteome predates the origin of vertebrate species that have anatomically large nervous systems. The expression patterns of synaptic proteins in mouse brain showed that each region expressed a similar set of postsynaptic proteins, although with different combinations of expression levels (**Fig. 4b**). The genes contributing most to anatomical variation in expression were typically of more recent origin and were substantially enriched with upstream signaling and structural components (reflecting their greater expansion in invertebrates and vertebrates). This preferential contribution to anatomical diversification in the mammalian brain made by synaptic proteome innovations of vertebrate origin suggests that tuning of expression of synaptic proteome genes resulted in functional diversification in brain regions.

Evolution of synaptic signaling complexity

How might the increased complexity of postsynaptic signaling complexes in vertebrates compared with invertebrates influence the differences in vertebrate plasticity and behavior? First, the expansion of upstream proteins, including receptors, provides a wider range of

specificity elements for ligands and extracellular signals. Second, the combinations and organization of neurotransmitter receptor complexes show substantial differences. In particular the NMDA receptor of invertebrates has a single NR2 subunit and a single Dlg MAGUK adaptor, thereby forming a single complex. In contrast, vertebrates have four of each, allowing for 16 potential NR2-Dlg complexes. Mutations in mice support a model in which the increased molecular complexity of mammalian MASCs contribute to diversity in behavioral and electrophysiological signaling. Comparison of Dlg2 (PSD-93), Dlg3 (SAP102) and Dlg4 (PSD-95) mutants shows unique synaptic plasticity phenotypes in response to different patterns of neural stimuli in the CA3-CA1 synapses of the hippocampus. At the behavioral level, there are distinct cognitive phenotypes in learning tasks for these mutants[4,6]. Similarly, the distinct mammalian NR2 subunits have specific signaling and behavioral phenotypes[7,34,35], as do many (>40) other MASC proteins drawn from various functional protein classes[16] (http://www.genes2cognition.org/db/).

Complexity differences between invertebrate and vertebrate MASCs are accompanied by differences in interaction domains that are responsible for organizing the signaling complexes[36]. The vertebrate NR2 cytoplasmic domain, which is responsible for binding MAGUK proteins (via the terminal PDZ interaction motif) and multiple postsynaptic signaling molecules, consists of ~600 residues[37–40]. In contrast, the invertebrate NR2 subunit possesses a short ~100-residue domain that lacks most of these interaction motifs, but retains the PDZ motif[36]. These differences in the organization and complexity of signaling complexes would be expected to result in postsynaptic signaling networks with distinct computational capabilities between vertebrates and invertebrates[16].

Anatomical diversification and evolution of the synapse proteome

Comparison of phylogeny and brain expression revealed unexpected relationships; recently evolved genes encoding upstream signaling and structural components of pathways contribute most to anatomical diversity (**Fig. 4b**). Electron microscopy studies of single synapses in rodents reveal that this diversity (for NR2 and MAGUK proteins) distinguishes individual synapses[41,42]. Although the full extent of synaptic diversity is unknown, we observed that the differential anatomical expression patterns of proteins in MASCs produces a signature or expression bar code. Given the large number of postsynaptic components and their variation in levels, there are very large numbers of potential combinations of complexes or synapses. The variation in levels of expression in brain regions was probably driven by mutations in cis-regulatory sequences of expanded vertebrate gene families, a process known as subfunctionalization. Both the comparison of control elements in these synaptic genes and comparative proteomics between brain regions in different species should provide further insight into the mechanisms generating synapse diversity.

Our model of synapse evolution indicates an ancestral or prototype synapse, which has been elaborated on to provide species with a mechanism for anatomical diversification and specialization of synapse function. It follows that synaptic proteome complexity contributes to the computational and cognitive properties of the brain and should be considered, along with differences in neuron number and connectivity, when interpreting evolutionary differences in behavior. This substrate for synapse diversification leads to the speculation that synaptic molecular complexity may have been a prerequisite for anatomical and functional changes underlying the emergence of complex behavioral repertoires. Although the complexity and diversity of human synaptic proteomes remains to be explored, it should be noted that both ancient (for example, NF1) and recent components (for example, Dlg3) of MASCs are encoded by genes that are responsible for heritable cognitive impairments. Moreover, some MASC components (for example, NR2A) show evidence of higher rates of evolution in primate compared with rodent lineages[43]. The synaptic proteome datasets from invertebrates and vertebrates provide a new approach for studying nervous system evolution and diversification and may be used to shed light on the origins of complex behaviors.

METHODS

Isolation of *Drosophila* MASCs. To prepare the affinity resin, we washed 0.5 ml of activated support Affi-Gel 10 (Bio-Rad) with 20 ml of chilled distilled water and 5 ml of chilled MOPS solution (50 mM MOPS, pH 8.0, and 50 mM CaCl2). We mixed 5 mg of peptide in 1 ml of MOPS solution, then mixed in 0.5 ml Affi-Gel 10 and coupled the peptide and the affinity resin for 4 h at 4 °C. The peptide was then blocked with 1 M Tris pH 9.0 at a 2:1 vol/vol ratio for 18 h at 4 °C, washed with 20 mM Tris pH 9.0 and stored at 4 °C. For affinity purification, 1 g of *D. melanogaster* heads were homogenized in 24 ml of lysis buffer (50 mM Tris, pH 7.4, 0.5 % Nonidet P-40 (vol/vol, Roche Diagnostics), 50 mM NaF, 20 μM $ZnCl_2$, 1 mM o-vanadate, 1mM PMSF, 2 ug ml^{-1} aprotinin and 2 ug ml^{-1} leupeptin) using an OMNI 2000 homogenizer (Omni International) on ice. Sample was left in ice for 1 h and centrifuged using a JA25.50 rotor (Beckman) at 48,384*g* for 30 min at 4 °C, and the supernatant was cleared through a 5-μM filter. Extracts were mixed with affinity resins at a 100:1 (vol/vol) and incubated for 18 h at 4 °C with agitation. Resin was washed with 300 column volumes of lysis buffer lacking PMSF. Resin was mixed with an equivalent volume of elution buffer (5 mg ml^{-1} solution of the same peptide contained on the resin in washing buffer with pH adjusted to 9) and incubated for 2 h at 4 °C with agitation. The mixture of resin and elution buffer was centrifuged for 1 min at 956*g* rpm in an Eppendorf table centrifuge and the supernatant containing the eluted sample was recovered.

Detection of orthologs across species. Swiss-Prot and TrEMBL accessions of PSD, NRC and MASC proteins previously identified by mass spectrometry and immunoprecipitation[8] were mapped to human Ensembl genes from NCBI Build 35 using EnsMart[44]. Orthologs of these were identified in 19 species using the Ensembl Compara database (http://www.ensembl.org, Ensembl version 36). A full list of gene builds used is given in the **Supplementary Methods**. Pfam domains[45] and Gene Ontology terms (http://www.geneontology.org) for each ortholog were obtained via EnsMart. To infer species phylogeny, protein sequences of GAPDH were aligned using muscle[46]. The unrooted maximum likelihood tree was generated by PhyML[47].

Identification of fly MASC orthologs. Orthologs of the 220 fMASC genes, together with their orthology type, were retrieved from the Ensembl 46 database via EnsMart. To identify fMASC genes preceding the divergence of fly and mouse lineages, we identified orthologs in yeast and the 14 chordate species. A full list of gene builds used is given in the **Supplementary Methods**.

Statistical analysis. To assess the similarity in composition between the synaptic function subsets of fMASCs (67 proteins) and mMASCs/mPSDs (155 and 355, respectively), we calculated the probability of a random set of 67 proteins from mMASC or mPSD having the same number from each functional family as fMASC. Consider a set of N proteins, of which $n(i)$ belong to functional family $f(i)$ ($i = 1 \ldots n_f$). We wished to calculate the probability that a random selection of M proteins will contain $m(i)$ in functional family $f(i)$ ($i = 1 \ldots n_f$). There are $\rho[M] = N!/[M!(N-M)!]$ ways of selecting M proteins from N. Of these, $\rho[m(1), m(2), \ldots m(n_f)] = \Pi_i n(i)!/[m(i)![n(i) - m(i)]]!$ contain $m(i)$ in functional family $f(i)$ ($i = 1 \ldots n_f$). The probability is thus $\rho[m(1), m(2), \ldots m(n_f)]/\rho[M]$.

The statistical significance of an overlap between two sets of molecules was calculated using a previously described method[16]. We used this method to analyze several datasets: phylogenetic expansion of mouse MASC/PSD functional classes, phylogenetic expansion of mouse fMASC functional classes, the relationship between functional classes of mMASC/mPSD orthologs of fMASC components and orthology type, expression variability and phylogeny, and expression variability and mouse MASC/PSD functional classes. Further details are given in the **Supplementary Methods**.

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

ACKNOWLEDGMENTS

We thank L. Hansen and L. Valor for assistance with array data, M. Marshall and K. Page for literature curation, A.J. Vilella, members of Genes to Cognition program for discussions and J.V. Turner for editorial assistance. This work was supported by the Wellcome Trust (C.N.G.A., J.D.A., R.D.E., M.O.C., M.D.R.C., J.S.C. and S.G.N.G.), the Medical Research Council (A.J.P. and R.D.E.), GlaxoSmithKline (B.R.M.), Edinburgh University and the e-Science Institute (J.D.A.), and the European Molecular Biology Organization (A.B.).

AUTHOR CONTRIBUTIONS

R.D.E. conducted bioinformatic analysis of MASCs/PSDs and fMASCs. A.J.P. carried out bioinformatic and statistical analyses of MASCs/PSDs, fMASCs and expression datasets. C.N.G.A. and C.A.V. performed RNA and protein expression studies. B.R.M., A.B., M.O.C. and J.S.C. conducted fly proteomic studies. M.D.R.C. provided informatic support. R.D.E., A.J.P., C.G.N.A., J.D.A. and S.G.N.G. interpreted the results and prepared the manuscript. S.G.N.G. conceived and supervised the project.

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

1. Kandel, E.R. The molecular biology of memory storage: a dialogue between genes and synapses. *Science* **294**, 1030–1038 (2001).
2. Moore, B.R. The evolution of learning. *Biol. Rev. Camb. Philos. Soc.* **79**, 301–335 (2004).
3. Roth, G. & Dicke, U. Evolution of the brain and intelligence. *Trends Cogn. Sci.* **9**, 250–257 (2005).
4. Cuthbert, P.C. *et al.* Synapse-associated protein 102/dlgh3 couples the NMDA receptor to specific plasticity pathways and learning strategies. *J. Neurosci.* **27**, 2673–2682 (2007).
5. Komiyama, N.H. *et al.* SynGAP regulates ERK/MAPK signaling, synaptic plasticity and learning in the complex with postsynaptic density 95 and NMDA receptor. *J. Neurosci.* **22**, 9721–9732 (2002).
6. Migaud, M. *et al.* Enhanced long-term potentiation and impaired learning in mice with mutant postsynaptic density–95 protein. *Nature* **396**, 433–439 (1998).
7. Sprengel, R. *et al.* Importance of the intracellular domain of NR2 subunits for NMDA receptor function *in vivo*. *Cell* **92**, 279–289 (1998).
8. Collins, M.O. *et al.* Molecular characterization and comparison of the components and multiprotein complexes in the postsynaptic proteome. *J. Neurochem.* **97** Suppl 1, 16–23 (2006).
9. Dosemeci, A. *et al.* Composition of the synaptic PSD-95 complex. *Mol. Cell Proteomics* **6**, 1749–1760 (2007).
10. Husi, H., Ward, M.A., Choudhary, J.S., Blackstock, W.P. & Grant, S.G. Proteomic analysis of NMDA receptor–adhesion protein signaling complexes. *Nat. Neurosci.* **3**, 661–669 (2000).
11. Collins, M.O. *et al.* Proteomic analysis of *in vivo* phosphorylated synaptic proteins. *J. Biol. Chem.* **280**, 5972–5982 (2005).
12. Yamauchi, T. Molecular constituents and phosphorylation-dependent regulation of the postsynaptic density. *Mass Spectrom. Rev.* **21**, 266–286 (2002).
13. Yoshimura, Y. *et al.* Molecular constituents of the postsynaptic density fraction revealed by proteomic analysis using multidimensional liquid chromatography-tandem mass spectrometry. *J. Neurochem.* **88**, 759–768 (2004).
14. Grant, S.G., Marshall, M.C., Page, K.L., Cumiskey, M.A. & Armstrong, J.D. Synapse proteomics of multiprotein complexes: en route from genes to nervous system diseases. *Hum. Mol. Genet.* **14** Spec No. 2, R225–234 (2005).
15. Laumonnier, F., Cuthbert, P.C. & Grant, S.G. The role of neuronal complexes in human X-linked brain diseases. *Am. J. Hum. Genet.* **80**, 205–220 (2007).
16. Pocklington, A.J., Cumiskey, M., Armstrong, J.D. & Grant, S.G.N. The proteomes of neurotransmitter receptor complexes form modular networks with distributed functionality underlying plasticity and behavior. *Mol. Syst. Biol.* **2**, E1–E14 (2006).
17. Delsuc, F., Brinkmann, H., Chourrout, D. & Philippe, H. Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* **439**, 965–968 (2006).
18. Cyert, M.S. Genetic analysis of calmodulin and its targets in *Saccharomyces cerevisiae*. *Annu. Rev. Genet.* **35**, 647–672 (2001).
19. Elion, E.A., Qi, M. & Chen, W. Signal transduction. Signaling specificity in yeast. *Science* **307**, 687–688 (2005).
20. Erdman, S. & Snyder, M. A filamentous growth response mediated by the yeast mating pathway. *Genetics* **159**, 919–928 (2001).
21. Harashima, T., Anderson, S., Yates, J.R., III & Heitman, J. The kelch proteins Gpb1 and Gpb2 inhibit Ras activity via association with the yeast RasGAP neurofibromin homologs Ira1 and Ira2. *Mol. Cell* **22**, 819–830 (2006).
22. Palecek, S.P., Parikh, A.S. & Kron, S.J. Sensing, signalling and integrating physical processes during *Saccharomyces cerevisiae* invasive and filamentous growth. *Microbiology* **148**, 893–907 (2002).
23. Nichols, S.A., Dirks, W., Pearse, J.S. & King, N. Early evolution of animal cell signaling and adhesion genes. *Proc. Natl. Acad. Sci. USA* **103**, 12451–12456 (2006).
24. Sakarya, O. *et al.* A postsynaptic scaffold at the origin of the animal kingdom. *PLoS ONE* **2**, e506 (2007).
25. te Velthuis, A.J., Admiraal, J.F. & Bagowski, C.P. Molecular evolution of the MAGUK family in metazoan genomes. *BMC Evol. Biol.* **7**, 129 (2007).
26. Yasuyama, K., Meinertzhagen, I.A. & Schurmann, F.W. Synaptic organization of the mushroom body calyx in *Drosophila melanogaster*. *J. Comp. Neurol.* **445**, 211–226 (2002).
27. Ruiz-Canada, C., Koh, Y.H., Budnik, V. & Tejedor, F.J. DLG differentially localizes Shaker K^+-channels in the central nervous system and retina of *Drosophila*. *J. Neurochem.* **82**, 1490–1501 (2002).
28. Husi, H. & Grant, S.G. Isolation of 2,000-kDa complexes of *N*-methyl-*D*-aspartate receptor and postsynaptic density–95 from mouse brain. *J. Neurochem.* **77**, 281–291 (2001).
29. Altschul, S.F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
30. Magdaleno, S. *et al.* BGEM: an in situ hybridization database of gene expression in the embryonic and adult mouse nervous system. *PLoS Biol.* **4**, e86 (2006).
31. Zapala, M.A. *et al.* Adult mouse brain gene expression patterns bear an embryologic imprint. *Proc. Natl. Acad. Sci. USA* **102**, 10357–10362 (2005).
32. Lein, E.S., Zhao, X. & Gage, F.H. Defining a molecular atlas of the hippocampus using DNA microarrays and high-throughput *in situ* hybridization. *J. Neurosci.* **24**, 3879–3889 (2004).
33. Toledo-Rodriguez, M. *et al.* Correlation maps allow neuronal electrical properties to be predicted from single-cell gene expression profiles in rat neocortex. *Cereb. Cortex* **14**, 1310–1327 (2004).
34. Kutsuwada, T. *et al.* Impairment of suckling response, trigeminal neuronal pattern formation and hippocampal LTD in NMDA receptor epsilon 2 subunit mutant mice. *Neuron* **16**, 333–344 (1996).
35. Sakimura, K. *et al.* Reduced hippocampal LTP and spatial learning in mice lacking NMDA receptor epsilon 1 subunit. *Nature* **373**, 151–155 (1995).
36. Ryan, T.J., Emes, R.D., Grant, S.G. & Komiyama, N.H. Evolution of NMDA receptor cytoplasmic interaction domains: implications for organization of synaptic signaling complexes. *BMC Neurosci.* **9**, 6 (2008).
37. Chung, H.J., Huang, Y.H., Lau, L.F. & Huganir, R.L. Regulation of the NMDA receptor complex and trafficking by activity-dependent phosphorylation of the NR2B subunit PDZ ligand. *J. Neurosci.* **24**, 10248–10259 (2004).
38. Li, B.S. *et al.* Regulation of NMDA receptors by cyclin-dependent kinase 5. *Proc. Natl. Acad. Sci. USA* **98**, 12742–12747 (2001).
39. Nakazawa, T. *et al.* NR2B tyrosine phosphorylation modulates fear learning as well as amygdaloid synaptic plasticity. *EMBO J.* **25**, 2867–2877 (2006).
40. Prybylowski, K. *et al.* The synaptic localization of NR2B-containing NMDA receptors is controlled by interactions with PDZ proteins and AP-2. *Neuron* **47**, 845–857 (2005).
41. Petralia, R.S., Sans, N., Wang, Y.X. & Wenthold, R.J. Ontogeny of postsynaptic density proteins at glutamatergic synapses. *Mol. Cell. Neurosci.* **29**, 436–452 (2005).
42. Sans, N. *et al.* A developmental change in NMDA receptor–associated proteins at hippocampal synapses. *J. Neurosci.* **20**, 1260–1271 (2000).
43. Dorus, S. *et al.* Accelerated evolution of nervous system genes in the origin of *Homo sapiens*. *Cell* **119**, 1027–1040 (2004).
44. Kasprzyk, A. *et al.* EnsMart: a generic system for fast and flexible access to biological data. *Genome Res.* **14**, 160–169 (2004).
45. Bateman, A. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **28**, 263–266 (2000).
46. Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
47. Guindon, S. & Gascuel, O. A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696–704 (2003).
48. White, J.G., Southgate, E., Thomson, J.N. & Brenner, S. The structure of the nervous system of the nematode *C. elegans*. *Phil. Trans. R. Soc. Lond. B* **314**, 1–340 (1986).
49. Truman, J.W., Taylor, B.J. & Awad, T.A. in *The Development of Drosophila melanogaster* (eds. Bate, M. & Martinez Arias, A.) 1245–1276 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1993).
50. Imai, J.H. & Meinertzhagen, I.A. Neurons of the ascidian larval nervous system in *Ciona intestinalis*. I. Central nervous system. *J. Comp. Neurol.* **501**, 316–334 (2007).