The synapse is fundamental to the structure and function of the brain through its role in connecting neurons into circuits. Experimental verification of all known synaptic proteins and their receptors, adhesion/cytoskeletal proteins, scaffold proteins, and neurotransmitters has greatly increased over the past decade, creating a need for a comprehensive resource to integrate information about synaptic genes and proteins (the synaptome). The synaptome is integral to the function of the brain and may be an important source of dysfunction underlying many neuropsychiatric disorders. Consequently, it is an excellent candidate for large-scale genomic and proteomic study. While the tools and databases available for the annotation of high-throughput DNA and protein data are generally robust, a comprehensive resource dedicated to the integration of information about the synapse is lacking.

Results: We present an integrated database, called SynaptomeDB, to retrieve and annotate genes comprising the synaptome. These genes encode components of the synapse including neurotransmitters and their receptors, adhesion/cytoskeletal proteins, scaffold proteins, membrane transporters. SynaptomeDB integrates various and complex data sources for synaptic genes and proteins.

Availability: http://psychiatry.jhmi.edu/SynaptomeDB/
Contact: mpiroz1@jhmi.edu
Supplementary information: Supplementary data are available at Bioinformatics online.

Received on September 30, 2011; revised on January 2, 2012; accepted on January 18, 2012

1 INTRODUCTION

The synapse is fundamental to the structure and function of the brain through its role in connecting neurons into circuits. As a result, the synapse is an excellent target of large-scale study of neuropsychiatric disorders. Over the past decade, the number of identified synaptic proteins has increased dramatically, creating a need for a comprehensive resource to integrate information about synaptic genes and proteins (the ‘synaptome’) from multiple heterogeneous sources. These genes encode components of the synapse including neurotransmitters and their receptors, adhesion/cytoskeletal proteins, scaffold proteins, membrane transporters and others. Here, we report on an integrated database, SynaptomeDB, which provides a detailed and experimentally verified annotation of all known synaptic proteins.
(Liberton et al., 2011) to identify pathways that are enriched for synaptic genes, which can inform subsequent biological analyses. Here, the proportion of genes in a given pathway appearing on the SynaptomeDB list is compared with the proportion of genes not appearing on the list, and a hypergeometric test (Holmans, 2010) is performed to test for differences in these proportions. This analysis is also fully automated and can be updated as new genes and sets are identified.

3 RESULTS

We assembled a list of genes (n = 1886) that encode all known proteins of the synapse. This comprises 575 genes encoding proteins in the presynaptic nerve terminal and active zone, 107 from the synaptic vesicles and 1755 from the postsynaptic density (there is some overlap between categories). The list includes strong candidates for a number of neuropsychiatric disorders such as, for example, ANK3 for bipolar disorder (Ferreira et al., 2008), GRM7 for major depression (Shyn et al., 2011), PDE4B for schizophrenia (Kahler et al., 2010) and SHANK3 for autism (Gauthier et al., 2009). SynaptomeDB is a database with a web front application that integrates the various and complex data sources for these synaptic genes.

3.1 Database design and features

The database is created using MySQL 5.5. The parsers are written in perl and Bioperl (Stajich et al., 2002). The Ensembl BioMart is also used to create some of the tables. A conceptual model of the database is shown in Supplementary Material 3. These tables describe fundamental information about a particular gene: name, description, associated accession numbers, chromosome location, function and comparative map information among other variables. Information from Ensembl also occupies a significant part of the database. It is important to note that no extensive cleaning of the data is performed during the database creation and update process. As detailed in Supplementary Material 1, the major cleaning process involves character screening to make sure the data is compatible for HTML viewing as well as database query. This allows automatic updates and eliminates some well-known problems created by data cleaning.

3.2 Web interface

SynaptomeDB provides a user-friendly web interface. Users can query SynaptomeDB using gene information such as names, gene IDs, synonyms and genomic regions. The output consists of a graphical representation of protein structure from PDB (Berman et al., 2003) and protein domain architecture from HPRD (Keshava et al., 2009). All information was hyperlinked to its original resources. SynaptomeDB allows the user to export multiple samples from different sample sets, in a desired order, to a number of common file formats including Excel, Word, CSV and XML. The web interface of SynaptomeDB provides a rich set of functions for searching the database. In general, search results are initially presented as the summary statements of individual gene records contained in SynaptomeDB, along with additional links to the gene detail page that reveal all details of the gene records returned by the query. A simple text search function is also provided to enable maximum flexibility in searching all records. The advanced search page provides complex searching functions. General database statistics are shown on the home page and reveal a quick summary of genes, as well as the last updates of the system.

4 FUTURE PLANS AND CONCLUSIONS

The database was constructed following guideline described previously (Kirov et al., 2005). It can be used to answer complex queries, such as defining a set of candidate genes based on the genome localization or specific function. The database provides a valuable resource to both experimental and bioinformatics groups by bringing together different sources of information and functional annotation in one place, and in a high-throughput fashion. A syndrome-based strategy for psychiatric genetic sequencing is valuable because there is evidence for synaptic proteins playing a role in psychiatric disorders (Glessner et al., 2010), and because these proteins represent the most 'drugable' targets for pursuit of novel therapies. Our application will further research in this area both in its current form and with additional modifications that will include incorporating navigation based on GO and functional pathways and networks among the Synaptome genes in the DB and to also include or link with protein-protein interactions. We intend to extend SynaptomeDB to connect to other psychiatric resources such as SZGene (Allen et al., 2008) and ALCgene (Bertram et al., 2007), and also to integrate variants from several ongoing studies that include synaptome genes, including the 1000 Genomes Project.

Funding: National Institutes of Health (R01-MH087970 to J.B.P., K01-MH093809 to M.P.); (R01-MH083738 to P.P.Z.); Johns Hopkins Brain Science Institute.

Conflict of Interest: none declared.

REFERENCES


