To be submitted as ‘NeuroResource’ article to Neuron (commissioning editor: Ted Dobie)

**Title** (max 12 words, 150 characters incl. spaces)**:**

**SynGO: an evidence-based, expert-curated knowledgebase for the synapse**

**Authors:** see page 2

**Correspondence:**

Guus Smit (guus.smit@cncr.vu.nl) and Matthijs Verhage (matthijs@cncr.vu.nl)

Center for Neurogenomics and Cognitive Research (CNCR)

VU University and VU Medical Center

De Boelelaan 1087

1081 HV Amsterdam, The Netherlands

Tel: +31 (0) 20 59 86936

**Keywords** (up to 10):synapse, gene annotation, gene ontology, synaptic plasticity, synaptopathies, enrichment study, gene-set analysis, synaptic proteome network, synaptome

**First authors:** Frank Koopmans[[1]](#footnote-1), Pim van Nierop[[2]](#footnote-2)

**Second authors:** Maria Andres-Alonso24, Andrea Byrnes26, Tony Cijsouw14, Marcello P. Coba[[3]](#footnote-3), L. Niels Cornelisse1, Ryan J. Farrell29, Hana L. Goldschmidt19, Daniel P. Howrigan26, Natasha K. Hussain19,20, Cordelia Imig15, Arthur P.H. de Jong22, Hwajin Jung33, Mahdokht Kohansalnodehi21, Barbara Kramarz[[4]](#footnote-4), Noa Lipstein15, Ruth Lovering5, Harold MacGillavry18, Huaiyu Mi33, Momchil Ninov21, David Osumi-Sutherland[[5]](#footnote-5), Rainer Pielot17, Karl-Heinz Smalla17, Haiming Tang33, Katherine Tashman26, Ruud F.G. Toonen1, Chiara Verpelli30, Rita Reig Viader13, Kyoko Watanabe28, Jan van Weering1

**Middle authors:** Tilmann Achsel12, Ghazaleh Ashrafi29, Nimra Asi29, Tyler C. Brown31,Pietro De Camilli[[6]](#footnote-6), Marc Feuermann[[7]](#footnote-7), Rebecca Foulger26, Pascale Gaudet7, Anouska Joglekar11, Alexandros Kanellopoulos12, Robert Malenka[[8]](#footnote-8), Vittoria Mariano12, Camila Pulido29, Jaime de Juan Sanz29, Morgan Sheng[[9]](#footnote-9), Thomas C. Südhof[[10]](#footnote-10) Hagen U. Tilgner[[11]](#footnote-11)

**Senior authors:** Claudia Bagni[[12]](#footnote-12), Àlex Bayés[[13]](#footnote-13), Thomas Biederer[[14]](#footnote-14), Nils Brose[[15]](#footnote-15), John Chua[[16]](#footnote-16), Daniela C. Dieterich17, Eckart D. Gundelfinger[[17]](#footnote-17),Casper Hoogenraad[[18]](#footnote-18), Richard L. Huganir[[19]](#footnote-19)[[20]](#footnote-20), Reinhard Jahn[[21]](#footnote-21), Pascal S. Kaeser[[22]](#footnote-22), Eunjoon Kim[[23]](#footnote-23), Michael R. Kreutz[[24]](#footnote-24), Peter S. McPherson[[25]](#footnote-25), Ben M. Neale[[26]](#footnote-26), Vincent O’Connor[[27]](#footnote-27), Danielle Posthuma[[28]](#footnote-28), Timothy A. Ryan[[29]](#footnote-29), Carlo Sala[[30]](#footnote-30)

**Last authors:** Guoping Feng[[31]](#footnote-31), Steven E. Hyman[[32]](#footnote-32), Paul D. Thomas[[33]](#footnote-33), August B. Smit2, Matthijs Verhage1

## SUMMARY (150 words max, current: 149)

Synapses are fundamental computational units of the brain and synaptic dysregulation is central to many brain disorders (‘synaptopathies’). However, systematic annotation of synaptic genes and ontology of synaptic processes are currently lacking. We have established SynGO, an interactive knowledgebase that synthesizes available research about synapse biology using Gene Ontology (GO) annotations to novel ontology terms for 87 synaptic locations and 179 synaptic processes. SynGO annotations are exclusively based on published, expert-curated evidence. Using 2922 annotations for 1112 genes, we show that synaptic genes are exceptionally well conserved and less tolerant to mutations than other (brain-expressed) genes. Many SynGO terms are significantly overrepresented among common gene variation associated with intelligence, educational attainment, ADHD, autism and bipolar disorder and among *de novo* variants associated with neurodevelopmental disorders including schizophrenia. Taken together, SynGO is a public, universal reference for synapse research and online analysis for gene enrichment studies (<https://syngoportal.org> and <http://geneontology.org>).

## INTRODUCTION

Synapses are computational units of the brain that provide the foundation for higher level information integration in dendritic segments, whole neurons and neuronal networks. Synaptic transmission and use-dependent changes in synaptic strength (synaptic plasticity) are now firmly established as the main underlying principles of cognitive processes, such as memory formation and retrieval, perception, sensory processing, attention, associative learning, and decision making ([Abdou et al., 2018](#_ENREF_1); [Groschner et al., 2018](#_ENREF_28); [Kandel, 2001](#_ENREF_33); [Petersen and Crochet, 2013](#_ENREF_49); [Ripolles et al., 2018](#_ENREF_53)). Synaptic dysregulation is generally recognized as a central underlying (causal) factor in many brain disorders (termed ‘synaptopathies’ ([Boda et al., 2010](#_ENREF_10); [Bourgeron, 2015](#_ENREF_11); [Grant, 2012](#_ENREF_27); [Monday and Castillo, 2017](#_ENREF_43))), such as autism spectrum disorder (ASD), ADHD, schizophrenia, Alzheimer’s disease and Parkinson’s disease ([Arnsten et al., 2012](#_ENREF_3); [Bourgeron, 2015](#_ENREF_11); [De Rubeis et al., 2014](#_ENREF_19); [Fromer et al., 2014](#_ENREF_23); [Heutink and Verhage, 2012](#_ENREF_30); [Selkoe, 2002](#_ENREF_58); [Soukup et al., 2018](#_ENREF_62); [Spires-Jones and Hyman, 2014](#_ENREF_63); [Sudhof, 2008](#_ENREF_64)). Despite these intense investigations and a large variety of research on synaptic proteins, their subcellular organization and specific functions, only sparse efforts have been made to establish systematic resources for synapse biology and pathobiology studies. In particular, the ontology of synaptic processes is poorly defined, which prevented the systematic annotation of synaptic proteins/genes.

The Gene Ontology (GO) is the most widely used resource for gene annotations. The resource has two components: the ontology, a computational structure that defines a framework of possible terms to describe gene functions and their relationships, and GO annotations, statements linking a specific gene to a specific term ([Ashburner et al., 2000](#_ENREF_4); [The Gene Ontology, 2018](#_ENREF_67)). The ontology is divided into three aspects: (i) molecular function (MF), defining the molecular activities of gene products (e.g. *protein kinase activity*); (ii) cellular component (CC), defining where gene products are active; and (iii) biological process (BP), defining the process that gene products carry out, usually together with multiple gene products. Relationships between CC terms generally specify how smaller structures are parts of larger ones and relationships between BP terms how sub-processes contribute to larger ones. The accuracy of GO annotations depends on (i) how well the ontology represents MF, CC and BP terms for given systems, i.e., synapses; and (ii) how well experimental evidence is converted into annotations.

Using existing annotations to synaptic GO terms and synaptic gene sets, several studies have shown that synaptic genes are significantly enriched in genetic variation associated with a variety of brain traits ([Savage et al., 2018](#_ENREF_57); [Zwir et al., 2018](#_ENREF_74)) and have produced valuable leads to understand the role of synapse function and dysfunction for these ([De Rubeis et al., 2014](#_ENREF_19); [Fromer et al., 2014](#_ENREF_23); [Mattheisen et al., 2015](#_ENREF_40); [Pedroso et al., 2012](#_ENREF_48); [Thapar et al., 2016](#_ENREF_66)). However, it is evident that the lack of systematic annotation of the genes encoding synaptic proteins also limits progress in this field. Current knowledge resources, including GO, have only limited representations of synapse biology, and lack a comprehensive ontology of synaptic processes and/or subcellular locations in the synapse. Rather than capturing the complexity of the current view of the synapse, existing resources are biased by uneven and patchy coverage of different aspects of synapse biology. Moreover, current resources include data not curated by synapse experts and a large fraction of the data in synaptic resources has been recruited in an unsupervised manner, for instance by automated text mining, or by large-scale experimentation with noteworthy false-positive rates, such as bulk proteomics analyses and yeast two-hybrid studies. Thresholds for inclusion are not systematically defined and are typically low. Together these limitations often engender incorrect conclusions, for instance in studies regarding links between synapses and brain traits.

To resolve these limitations, we established SynGO, a partnership between the GO Consortium and 15 synapse expert laboratories in Europe, North America and Asia, for the systematic annotation of synaptic proteins. SynGO experts have developed an extensive ontology to represent synaptic locations (87 terms) and synaptic processes (179 terms) and generated almost 3000 annotations of synaptic genes/proteins to these terms, based on a novel comprehensive evidence tracking system that classifies evidence according to experiment types, model systems and target engagement types (gene modifications, antibody binding etc.), using only published data sets. Using SynGO, we observed that synaptic genes are exceptionally well conserved, intolerant to mutations and enriched for gene sets associated with many brain traits, such as IQ and educational attainment, and for brain disorders such as ASD, ADHD and bipolar disorder. SynGO provides a unique, publicly accessible knowledgebase ([www.syngoportal.org](http://www.syngoportal.org)) as a universal reference for synapse research and education, and for enrichment studies on genomic associations, mRNA profiling and proteomic data.

## RESULTS

## SynGO ontologies for synaptic locations & processes provide comprehensive frameworks for synaptic gene annotation

## To systematically annotate synaptic genes, we designed a generic synapse model as a conceptual starting point, defining locations at the synapse and processes related to the synapse and refined this model iteratively until consensus was reached among expert labs worldwide (Fig 1). Subsequently, we created GO terms for cellular compartments (CC) and biological processes (BP) for synapses and defined their relationships. At the top level of the CC hierarchy (Fig 2A), synaptic proteins can be localized to the presynapse, the postsynapse, the synaptic cleft, the extra-synaptic space and synaptic membranes (the latter term is used when no distinction is possible between pre- and postsynaptic). From these high-level terms, up to 4 additional hierarchical levels were defined for pre- or postsynaptic cytosol or membrane, or organelles within these compartments. SynGO CC ontology added substantial precision to preexisting GO ontologies that contained 13 terms directly connected to the central ‘synapse’ term. SynGO maintained two terms (Fig. 2A, green symbols) and made 11 obsolete (Fig. 2A, purple symbols). Some of these were replaced by similar but less composite terms and others were replaced with more specific terms further down in the hierarchical SynGO ontology. As a result, 142 SynGO CC ontologies were designed for accurate annotation of synaptic localizations. To visualize these elaborate ontologies and provide a standardized visualization of SynGO annotations, all CC terms populated with gene annotations in SynGO v1.0 (92/142 terms) were plotted in a circular fashion with the highest hierarchical term (synapse) in the center and each layer of child terms in outward concentric rings (Fig. 2C, see Table S2 for all term names). SynGO did not define mitochondria as part of CC, as mitochondrial proteins are already well annotated ([Calvo et al., 2016](#_ENREF_13); [Smith and Robinson, 2018](#_ENREF_61)).

## BP terms for synaptic processes and their relationships were also defined consistent with existing GO-terms, with pre- and postsynaptic processes, synaptic organization, synaptic signaling, axonal/dendritic transport, and metabolism as main terms, with downstream up to 5 levels of sub-classes (Fig 2B). The precision gain is again substantial for SynGO BP ontologies. In total, the BP ontology features 256 terms of which 212 are new. 192 of these BP ontology terms were populated with gene annotations in SynGO v1.0 and visualized in a sunburst plot (Fig 2D, analogous to Fig 2B, see Table S2 for all term names). Hence, this major innovation in CC and BP terms and their relationships provide substantially increased precision in synapse ontologies that together provide a comprehensive structure for the systematic annotation of synaptic genes/proteins and for future computational models and machine learning approaches to build better *in silico* functional tests of synaptic processes and dysfunction.

## SynGO is based on expert annotation and systematic evidence tracking

Current synaptic protein lists contain many unsupervised inclusions, in particular data from large-scale, automated experiments known to have substantial false positive rates. SynGO established a systematic evidence tracking protocol and annotation by synapse experts only, based exclusively on published experimental data (PubMed). The SynGO annotation workflow (Fig. S1) was implemented in a web-interface and used by synapse experts to assign synaptic localizations (CC) or functions (BP). For both CC and BP evidence, classifications were designed for the model system(s) used (Fig. S2, e.g., intact tissue, cultured neurons, etc). For synaptic localization (CC), microscopy and biochemical studies were defined as the main experimental classes, each with several sub-classes. For functional studies, experimental classes were defined based on perturbation type (genetic or pharmacological) and the methodology (assay) used to detect the consequences (e.g., microscopy, electrophysiology etc., again with several sub-classes). These classifications were made coherent with the Evidence and Conclusions Ontology (ECO) ([Giglio et al., 2018](#_ENREF_26)), and new ECO terms were defined. Together, these three dimensions of evidence, (i) model system/preparation, (ii) experimental perturbation and (iii) assay, provide a systematic, coherent and detailed definition of the evidence to annotate synaptic genes.

These three dimensions of evidence tracking were used to annotate individual synaptic proteins/genes to specific CC or BP terms in the SynGO ontologies (Fig. S2, Table S3). Detailed reference to these evidence types was stored as part of the annotation (PubMed ID, figure numbers, panels, see Table S3). Thus, with these specific links to published evidence (PubMed), experts provided a detailed rationale for each annotation, which can be reviewed by SynGO users. For any given study, annotations were made for the species used and these were subsequently mapped to the consensus human ortholog using HUGO Gene Nomenclature Committee (HGNC) data resources ([Yates et al., 2017](#_ENREF_72)). Annotations for orthologous genes in different species were possible and encouraged, yielding multiple annotations to the same consensus human ortholog originating from different species. In addition, we applied SynGO annotations in GO Phylogenetic Annotation ([Gaudet et al., 2011](#_ENREF_25)) to infer annotations to evolutionarily-related genes, using the experimentally-supported SynGO annotations as evidence. In this process, an expert biocurator reviewed all experimentally-supported GO annotations for all members of a gene family in >100 species in the context of a phylogenetic tree and inferred functions of experimentally uncharacterized genes in tens of other organisms. In the current SynGO 1.0 we did not systematically annotate different splice forms of single genes, because systematic evidence for splice site-specific subcellular localizations or functions is currently sparse. In cases where studies used different approaches to reach the same conclusion, multiple annotations for the same gene/protein to the same CC or BP terms were made frequently and were encouraged. Similarly, when evidence existed for annotating a single gene/protein to multiple CC or BP terms (multiple locations or functions) multiple annotations were made and encouraged. Following standard GO annotation practice, the same gene/protein may be annotated at different levels along the SynGO hierarchical ontology tree. For instance, initial evidence may indicate that a protein is involved in synaptic transmission (SynGO term *chemical synaptic transmission*; GO:0007268), a subsequent study may reveal the protein regulates presynaptic secretion (SynGO term *synaptic vesicle exocytosis*; GO:0016079) and the most recent study showing the protein regulates vesicle priming (SynGO term *synaptic vesicle priming*; GO:0016082).

Annotations completed by expert laboratories first passed through a quality control pipeline by the SynGO support team (Fig. S1) and were then added either directly to the SynGO database (<https://syngoportal.org>) or returned to the annotator teams if further editing was required. These annotations were also deposited in the Gene Ontology annotation repository (<http://geneontology.org>) as GO-CAM ontologies ([The Gene Ontology, 2018](#_ENREF_67)). GO-CAM is an extension of the standard GO annotation format that allows more expressive annotations, e.g. specifying the cell type using Cell Ontology terms ([The Gene Ontology, 2018](#_ENREF_67)), and multiple pieces of evidence for a single annotation. Together, this evidence tracking system, including detailed reference to the evidence (PMID, figure, panel), provides an excellent framework for comprehensive annotation of synaptic genes.

## SynGO 1.0 provides 2922 expert-curated annotations on 1112 synaptic genes

## Using these three dimensions of evidence tracking, 2922 expert-curated annotations were generated using candidate synaptic gene lists from published ([Lips et al., 2012](#_ENREF_38); [Ruano et al., 2010](#_ENREF_56)) and unpublished data resources (EU-funded projects EUROSPIN and SYNSYS, see acknowledgements), proteomic data and input from expert labs. The annotations were subjected to quality control and, typically after iterative optimization, deposited in the SynGO database and the central Gene Ontology knowledgebase ([The Gene Ontology, 2018](#_ENREF_67)), see Fig S1. In total, we found compelling evidence for 1112 unique synaptic genes/proteins. These were admitted to the SynGO 1.0 knowledgebase. For most genes/proteins, both subcellular localization (CC) and biological process (BP) evidence was found (60%, Fig 3A), for the remaining 40%, evidence was lacking for either CC or BP and only one term was included. A core set of synaptic proteins was annotated to ≥3 CC or BP terms (Fig. 3B). Most evidence was obtained from studies of rodent species (Fig. 3C) of either intact tissue or cultured neurons (Fig. 3D). Microscopy and biochemical fractionation were the two main assay types used to make CC annotations, whereas BP annotations were based on a larger array of assay types assessing synaptic function (Fig. 3E). Together, these 2922 expert-curated annotations on 1112 synaptic genes/proteins, with a core set annotated to ≥3 CC or BP terms, provide an excellent annotation collection for descriptive studies, functional analyses of synaptic genes and gene enrichment studies.

**The structure of synaptic genes is very different from other genes**

As a first descriptive analysis, we compared basic structural features of SynGO-annotated synaptic genes with other genes. Human gene features were extracted from BioMart (GRCh38.p12) and Ensembl web services. Interestingly, synaptic genes were found to be different from other (non-SynGO) genes in many respects. Synaptic genes were on average more than twice as long as other genes (257% of non-SynGO genes, Fig. 4A), with 1.6 fold longer cDNA (Fig 4B). The number of known protein coding transcripts was 1.7 fold higher (Fig 4C) and the sequence of introns + exons (immature transcript length) for protein coding transcripts was more than 2 fold longer (Fig 4D). Protein coding transcripts for synaptic genes also contained 1.4 fold more introns (Fig. 4E) and these were 1.7 fold longer (Fig. 4F).

To compare SynGO genes to other brain-expressed genes, we defined two control gene sets: (A) *brain-enriched genes*: 6600 genes with the most *brain-enriched* expression patterns, i.e., maximal expression difference between brain and other tissues ([Ganna et al., 2016](#_ENREF_24)); and (B) ‘top N’ genes most highly expressed in brain, with N equal to the number of unique genes in the SynGO set (1112). For comparisons of gene size, introns, cDNA length, differences between SynGO genes and control sets A and B were generally smaller, but still highly significant (Fig. S3A-L). We also investigated the complexity of isoform expression of synaptic genes within cerebellar neurons using recently published full-length RNA sequencing data ([Gupta et al., 2018](#_ENREF_29)). Synaptic genes expressed a higher number of distinct isoforms, as compared to non-SynGO genes, per equal read counts than non-synaptic genes (Fig. S4).

We also analysed the number of posttranslational modifications, as important determinants of cell signalling, by testing the number of experimentally verified modifications obtained from dbPTM ([Huang et al., 2016](#_ENREF_31)) and Uniprot ([Consortium, 2018](#_ENREF_17)) per gene/protein and per amino acid (to correct for difference in average protein length; Fig. S5). The incidence of all major modifications, phosphorylation, ubiquitination, acetylation and S-nitrosylation appear to be all significantly higher in synaptic genes as compared to other genes. However, these observations might emerge, at least in part, from the fact that synaptic genes are more intensively studied experimentally.

**Synaptic genes emerged earlier in evolution than other genes, primarily in three major waves**

The evolution of synaptic genes follows a pattern that differs substantially from the overall pattern for all human genes (Fig 5A). Specifically, synaptic genes evolved primarily in three “waves” of innovation, during which modern-day synaptic genes were gained at a faster rate than other human genes. The first wave of emergence of SynGO genes, was prior to the last eukaryotic common ancestor (LECA), approximately 1800 million years (Mya) ([Kumar et al., 2017](#_ENREF_35)). While LECA was unicellular and obviously did not form synapses, it did possess cellular machinery that would later be co-opted for the synapse, such as vesicle trafficking, exocytosis and signal reception. The second wave was prior to the last common ancestor of the eumetazoa (multicellular animals), and corresponds with the first appearance of the synapse. The third wave was prior to the last common ancestor of vertebrates, suggesting significant synaptic evolution in this period. By this time, approximately 450 Mya, about 95% of all synapse genes were already in place, with very few additional synaptic genes appearing after that point. A similar trend, albeit with smaller differences, was observed when gene duplication events were not weighted (Fig S6). Figure 5B shows one of the few exceptions to this rule: additional genes have evolved due to recent duplication of the CPT1 gene, following the last common ancestor of placental mammals, resulting in an additional paralog placental mammals (CPT1C), whereas other mammals have only one two (CPT1A, CPT1B). Additional specification also occurred in zebrafish and frogs (Fig 5B) These data indicate that the synapse is highly conserved among modern vertebrates, as suggested before ([Emes et al., 2008](#_ENREF_21)), and suggest that 95% of the human synaptic genes in SynGO 1.0 are shared among all vertebrates.

**Synaptic gene expression is enriched in the brain**

We predicted that expression of SynGO genes is higher in the brain than in other tissues. To test this, we compared tissue specific expression using different gene-sets in GTEx v7 ([Consortium et al., 2017](#_ENREF_16)). Brain enrichment was computed by dividing the number of transcripts detected in brain over those in other tissues, expressed as log2 fold change (see Methods) and plotted against the expression level of this transcript in brain. As shown in Fig 6A, expression of SynGO genes is generally higher in brain than in other tissues, although some SynGO genes are in fact de-enriched in brain (below horizontal line at zero). SynGO genes that are highly expressed in brain have on average a similar enrichment factor as those expressed lower in the brain (Fig 6A-B).

We compared brain expression enrichment for different SynGO CC and BP terms. Several ontology terms within these ontologies, especially in BP, are predicted to be highly brain specific, e.g., trans-synaptic signaling, active zone assembly or postsynaptic density organization, whereas others are expected to be similar to terms outside the synapse and outside the brain, e.g., phosphatase and kinase pathways, such as protein kinase C and cAMP-dependent kinase. Indeed, specific analyses of individual SynGO terms in CC and BP ontologies revealed a large degree of heterogeneity among proteins annotated for different terms (Fig 6C-D). The pre- and postsynaptic plasma membranes and especially the postsynaptic density contain proteins that are highly significantly enriched in brain (Fig 6C). Active zones and synaptic vesicles, but not dense core vesicles, also contain significantly enriched proteins (Fig 6C). For BP, a strong enrichment was observed for most major synaptic processes except metabolism and transport (Fig 6D). Taken together, these data indicate that expression of SynGO genes is higher in brain than in other tissues, especially for locations/functions that are specific for synapses.

**Synaptic proteins are exceptionally intolerant to mutations**

The frequency of coding variants in the general population is an indication of the functional constraints. To test whether SynGO genes have the same loss-of-function mutation incidence as other genes, we used the probability of being loss-of-function intolerant (pLI) obtained from the Exome Aggregation Consortium (ExAC, ([Karczewski et al., 2017](#_ENREF_34)). The probability of being loss-of-function intolerant (pLI) was compared between all SynGO genes and control sets. A major difference in loss-of-function intolerance was observed; SynGO genes are exceptionally intolerant to loss-of-function mutations relative to non-SynGO, brain-enriched and ‘top N’ most highly brain expressed control genes (Fig. 7A-C). The distribution of a high pLI value was similar among CC and BP terms (Fig. 7D-E). In the CC ontology, pLI scores were particularly high (mean value ≥ 0.7) for PSD and active zone genes (which also contribute to parent terms). Interestingly, the synaptic vesicle and dense core vesicle annotated genes showed much lower pLI scores (mean value ≤ 0.5). Taken together, these data indicate that synaptic genes are exceptionally intolerant to loss-of-function mutations, indicating that functional constraints and evolutionary selection pressure on synaptic genes are much stronger than for other genes.

**Synaptic proteins in closely related SynGO terms are more likely to interact**

SynGO proteins annotated to the same ontology term or to closely related terms are predicted to be often in the same protein complexes or part of the same process and are thus more likely to interact. This prediction was tested using protein-protein interaction data available through StringDB v10.5 ([Jeanquartier et al., 2015](#_ENREF_32)), using the ‘high confidence’ interaction filter. Proteins reported to be in the same protein complexes were significantly overrepresented in synaptic genes annotated against the same CC term in SynGO (Fig. 8A) and also for the same BP term (Fig 8B). Hence, synaptic proteins annotated for the same CC or BP term are much more likely to interact and, vice versa, interacting synaptic proteins are much more likely to have the same localization or be part of a similar process.

**Different synaptic preparations contain largely overlapping synaptic protein collections**

SynGO enables the analysis of existing, large-scale proteomics data from biochemical preparations enriched for synaptic components, which may provide valuable leads contributing to future inclusion of new genes/proteins in synaptic data sets upon further experimentation. Therefore, we extracted data from 19 well-described and quantitative proteomic analyses available in literature for 3 biochemical preparations enriched for synaptic components: (A) synaptosome fractions (7 studies, ([Bayes et al., 2017](#_ENREF_6); [Biesemann et al., 2014](#_ENREF_8); [Chang et al., 2015](#_ENREF_14); [Filiou et al., 2010](#_ENREF_22); [Moczulska et al., 2014](#_ENREF_42); [Pandya et al., 2017](#_ENREF_46); [Wilhelm et al., 2014](#_ENREF_69))); (B) postsynaptic density fractions (PSD, 6 studies, ([Bayes et al., 2012](#_ENREF_5); [Bayes et al., 2017](#_ENREF_6); [Bayes et al., 2011](#_ENREF_7); [Collins et al., 2006](#_ENREF_15); [Pandya et al., 2017](#_ENREF_46); [Roy et al., 2018](#_ENREF_55))) and (C) active zone or docked vesicle fractions (5 studies, ([Abul-Husn et al., 2009](#_ENREF_2); [Boyken et al., 2013](#_ENREF_12); [Morciano et al., 2009](#_ENREF_44); [Morciano et al., 2005](#_ENREF_45); [Phillips et al., 2005](#_ENREF_50))). Synaptosome studies have identified between 894 and 3331 proteins (Fig. 9A). These protein collections contained between 17 and 39% of the SynGO cellular component annotated proteins. Together, 80% of proteins with a SynGO cellular component annotation were detected in at least 1 of the synaptosome preparations. PSD analyses typically identified smaller numbers of components, up to 1207 ([Roy et al., 2018](#_ENREF_55)).

The consensus set of proteins identified in at least three proteomic datasets per compartment contains 2621 unique proteins for synaptosome, 791 for PSD and 88 for active zone. The PSD components showed a large degree of overlap, i.e. 90%, with the synaptosome consensus set with only 76 proteins exclusively identified in the PSD consensus set (Fig. 9B). 73% (1906 proteins) of the synaptosome consensus set is not found in the PSD consensus set, 78% (2033 proteins) is not found in SynGO and in total 61% (1596 proteins) of the synaptosome consensus set was not found in either PSD, active zone or the SynGO database.

Active zone preparations yielded even smaller numbers of proteins, maximally 249 (Fig. 9A). These protein collections contained between 35 and 62% of SynGO annotated proteins, slightly more than synaptosome and postsynaptic density percentages. A total of 2084 proteins currently lacking SynGO cellular component annotation were identified in at least three proteomics datasets of synaptosome, active zone or PSD subcellular fractions (Fig. 9B).

Taken together, these data indicate that SynGO aids in dissecting overlap and differences in large synaptic protein sets that were purified in different synaptic preparations. Many proteins identified in such fractions await experimental validation before they can be annotated to SynGO CC and BP terms.

**Synaptic genes are enriched among genes associated with various brain traits**

Over-representation studies in large-scale experimental case/control data (enrichment studies) gain power with a higher confidence definition of the gene sets used. We predicted that expert-curated, evidence-based SynGO genes show robust associations with experimental data on brain traits where synapse function/dysfunction plays a central role and that SynGO genes outperform existing synapse gene sets/lists. We tested this prediction on genome-wide association study (GWAS) data for three physiological traits, educational attainment (EA) ([Lee et al., 2018](#_ENREF_36)), Intelligence Quotient (IQ) ([Savage et al., 2018](#_ENREF_57)) and human height ([Wood et al., 2014](#_ENREF_70)), and for five brain disorders, ADHD ([Demontis et al., 2016](#_ENREF_20)), autism spectrum disorder (ASD) ([Robinson et al., 2016](#_ENREF_54)), schizophrenia ([Pardinas et al., 2018](#_ENREF_47); [Ripke et al., 2013](#_ENREF_52)), bipolar disorder ([Psychiatric, 2011](#_ENREF_51)) and major depression ([Wray et al., 2018](#_ENREF_71)). SynGO synaptic genes and previously annotated synaptic genes in GO were compared for these traits to three control gene sets: all other genes, other genes with similar brain-enriched expression and genes with similar (high) conservation. Two analysis methods were used, MAGMA ([de Leeuw et al., 2015](#_ENREF_18)) and linkage disequilibrium score (LDSC) regression analysis ([Zheng et al., 2017](#_ENREF_73)).

Fig 10A shows two typical enrichment analyses using MAGMA for one of these traits, autism spectrum disorder (ASD). We observed highly significant enrichment for the presynaptic active zone and the postsynaptic density (CC-terms), for presynaptic functions and synapse assembly (BP-terms), when these SynGO genes were compared to all other genes (Fig 10A). These associations remained significant, albeit typically less strongly, when SynGO genes were also conditioned on gene expression values (Fig 10A, dark colors), or conditioned on homology conservation scores (Fig S7A-D). Interestingly, one set of SynGO genes, the postsynaptic ribosome genes, was not significant when compared to all other genes, but became significant when compared to brain-expressed genes. Hence, enrichment analysis for SynGO genes in ASD GWAS data reveals new and highly significantly enrichment for pre- and postsynaptic compartments and presynaptic processes.

Similar analyses were performed for all other traits listed above (Fig 10B). Synaptic genes were significantly enriched among genes associated with educational attainment, especially postsynaptic localizations and processes. Five ontology terms were enriched among genes associated with IQ/Intelligence, but none were associated with human height. Furthermore, many ontologies were enriched among genes associated with ADHD, with especially locations and functions related to the presynaptic active zone and presynaptic assembly being highly enriched (Fig 10B). Finally, strong enrichments of both pre- and postsynaptic ontologies were observed among genes associated with ASD, and for postsynaptic processes among genes associated with bipolar disorder (Fig 10B). Fig 10B depicts MAGMA gene-set analysis with conditioning on gene expression values, very similar conclusions were reached when additionally conditioning on homology conservation scores and when LDSC regression analysis was used instead of MAGMA (Fig S7A-D).

Taken together, SynGO genes are strongly enriched in GWAS data for brain disease and physiological traits in which synapse (dys)function is central, with new links becoming manifest between ASD and the synapse; ADHD and presynaptic genes; educational attainment and postsynaptic processes and several other links between synaptic genes and bipolar disorder and IQ/Intelligence.

**Synaptic genes are enriched among *de novo* protein-coding variants for all brain disorders**

In addition to GWAS studies, exome sequence studies of *de novo* coding variation have recently become available, allowing us to perform enrichment studies in SynGO genes among all *de novo* coding variation detected from several brain disorder patient populations. We analyzed enrichment for protein truncating (PTV) and missense mutations in SynGO genes for 4 brain diseases: Developmental Delay (DD, 4293 trios), Intellectual Disability (ID, 971 trios), ASD (3982 trios) and Schizophrenia (SCZ, 1024 trios), with non-syndromic Congenital Heart Defect (CHD, 1487 trios) and unaffected siblings (UNAFF SIB, 2216 trios) as non-affected classes. PTV and missense mutations were filtered if they were present in the ExAC reference database ([Lek et al., 2016](#_ENREF_37)), and *de novo* enrichment in each group was compared against a mutation model that estimates the expected mutation rate among each gene set. SynGO gene enrichment was compared to previously annotated synaptic genes in GO and to *matched brain-enriched genes***:** control gene sets with similar brain enrichment/specificity and gene size exactly matching SynGO genes. SynGO genes were robustly enriched for all 4 disease classes (Fig 11A-B), most strongly for ID (>2 fold enriched), but also for DD (1.6 fold enriched), ASD (1.4 fold enriched) and SCZ (1.3 fold enriched). All these enrichments for SynGO genes were substantially stronger than for synaptic genes previously annotated in GO, especially for DD and ID (Fig. 11A). PTVs and missense mutations in SynGO genes were not enriched for CHD-NS and in unaffected siblings (Fig. 11A).

To test the distribution of these enrichments within SynGO ontology terms, we plotted the enrichment p-values for each term as false colour values in SynGO CC and BP ontologies (Fig. 11C-D, Table S7). Highly enriched protein sets were unevenly distributed among locations and processes. For subcellular locations (CC) the strongest associations were observed in postsynaptic density and active zone, together with pre- and post-synaptic plasma membrane terms (Fig 11C). For biological processes (BP), the strongest associations accumulated in presynaptic processes, especially synaptic vesicle exocytosis and generation of the presynaptic membrane potential, with further association in postsynaptic processes and synapse organization (Fig 11D). Together these data show that de novo PTVs and missense mutations in SynGO genes were strongly enriched for *de novo* PTV and missense variation in all four brain disorder populations. Importantly, SynGO genes show more robust enrichments than previous synaptic GO-genes.

## DISCUSSION

This study describes SynGO, the first comprehensive knowledgebase that provides an expert community consensus ontology of the synapse. The ontology and annotations accumulated in SynGO provide a comprehensive definition of synapses, new unique features of synapses, new links between synapses and brain disorders and excellent future perspectives as an up-to-date interactive community resource. We deliver proof of principle application of the SynGO ontology and annotations for the analyses of gene/protein properties, evolutionary conservation, mRNA expression, loss of function tolerance, protein-protein interaction, enrichment in GWAS data for physiological traits and brain disorders, and in rare *de novo* coding variation for neurodevelopmental disorders including schizophrenia.

**SynGO provides a major step forward in defining synapses**

Adequately defining a biological system like the synapse requires a coherent and logical definition of its components, their relationships and how biological functions emerge from these. SynGO ontology is the first ontology to provide such definitions coherently for the synapse. The SynGO ontology has defined 87 Cellular Component (CC) and 179 Biological Process (BP) terms, that were designed in consensus by expert laboratories worldwide. Previous models suffered from the lack of a coherent, top-down design of synapse-related ontology terms and relations. Consequently, many heterogeneous terms, both specific and general, were positioned directly under the master term ‘synapse’ (see Fig 2A-B).

Defining synapses adequately also requires the underlying annotations to be accurate and reliable. SynGO is exclusively based on published, expert-curated evidence and detailed classification of this evidence. This is a major innovation that provides accountability for decisions made by experts and allows for structured discussions and resolving annotation disputes, in particular in the web-based SynGO resource (www.syngoportal.org). Moreover, different types of evidence can now be integrated in statistical models in a differential manner. For instance, evidence that is considered very strong can be given a higher weight than evidence less so. Finally, providing evidence tracking tools to (future) expert contributors engages the synapse research community, ensuring that SynGO annotations are based on solid evidence. Hence, the new SynGO evidence tracking provides a fundamental step forward for annotation accuracy, transparency and expert-engagement and a solid basis for future refinements in a biology-driven overall synaptic ontology framework.

Using SynGO 1.0 annotations, we show that the SynGO ontology indeed defines the synapse adequately and provides a major improvement over existing models. We show that (i) SynGO genes are indeed more evolutionary conserved than other genes (Fig 5), as previously shown ([Emes et al., 2008](#_ENREF_21)), and (ii) that synaptic genes are indeed brain enriched, with brain-specific aspects of synapses particularly enriched, as opposed to generic aspects, like transport and metabolism (Fig 6). Furthermore, (iii) SynGO proteins documented to interact in published protein-protein interaction data are much more likely to be annotated to the same ontology terms (Fig 8). Finally, (iv) enrichment of synaptic genes among genes associated with all tested traits in GWAS data (Fig 10) and among rare variants causing neurodevelopmental disorders (Fig 11), is without exception stronger for SynGO genes than for gene-sets previously annotated to the synapse. Together these four groups of observations confirm that SynGO defines synapses adequately, consistent with previous findings, and consistently outperforms previous gene set resources used in enrichment studies.

While the definition of a synapse is now becoming accurate and reliable, the definition of synaptic genes remains precarious. No cellular compartment operates in isolation, and components move in and out. Since GO annotations for location (CC) and process (BP) are independent, genes that regulate synaptic function do not necessarily have to be located in the synapse. In principle, this opens the possibility of annotating for instance transcription factors that regulate expression of synaptic genes. SynGO 1.0 currently only lists few of these examples, but it will eventually be useful to include such genes in SynGO annotation. Such genes can be easily excluded for analysis by filtering for CC terms, i.e., only genes that have a confirmed synaptic location will be retained. Other regulatory aspects of synapse function may include proteins derived from the extracellular matrix, axon, dendrite or glia, which are not yet accommodated in SynGO 1.0.

Taken together, SynGO provides a comprehensive definition of the synapse with new, elaborate and consensus ontologies, accurate and transparent evidence tracking and close to 3000 validated annotations. SynGO is ready to serve as a universal reference in synapse biology and for enrichment studies using –omics data, but also to form the fundament for future computational models to help understand synaptic computation principles in the brain and their dysregulation in disease**.**

**SynGO discovers unique features of synaptic genes and new disease links**

In addition to adequately defining synapses, SynGO also allowed us to identify several novel features of synapses and synaptic genes/proteins. First, we show that synaptic genes are structurally very different from other genes (Fig 4). Strikingly, SynGO genes are structurally very different than other genes, also relative tobrain-enriched genes. Second, nearly all synaptic genes have evolved prior to the last common ancestor of all vertebrates, >450M years ago, much earlier than many other human genes (Fig 5). Third, synaptic genes are exceptionally intolerant to mutations (Fig 7). We find that synaptic genes have accumulated more coding and non-coding sequence, which may have served to expand their transcriptional regulatory repertoire and diversification of functions of the encoded proteins. Moreover, larger genes with more intron-exon boundaries may have given rise to more alternatively spliced variants; a prediction that may soon become validated with the introduction of new long-read RNA sequencing. Also, mechanisms of gene duplication and splicing have generated expansion of synaptic gene diversity. Interestingly, as synaptic genes are found highly intolerant to mutation this diversification must have come with incorporating new essential synaptic functions, such as in features of plasticity, contributing to accelerating computational capabilities of the brain during evolution.

Synaptic dysregulation is central to many brain disorders (‘synaptopathies’). SynGO analyses described here strengthen the links between synapses and many brain traits (Fig 10-11). Many SynGO CC and/or BP terms are enriched among genes associated with educational attainment, intelligence (IQ), ADHD, ASD and bipolar disorder. In particular, novel links were discovered between educational attainment and postsynaptic processes. Exceptionally strong novel links were identified between ADHD and both pre- and postsynaptic genes, between ASD and presynaptic genes (in addition to the well-known links to the PSD, see ([Bourgeron, 2015](#_ENREF_11))) and between bipolar disorder and postsynaptic genes. One informative achievement of SynGO analyses is that, due to detailed structure of the SynGO ontology, genetic risk for each disease was mapped to specific synaptic locations and processes. The mapping resolution to specific terms is currently limited by the small number of genes/proteins annotated in some sub-classes in levels 3 and down. More synapse research is necessary to drive this refinement to saturation and allow more specific and definitive associations between genetic risk for brain disorders and distinct synaptic locations and processes.

**SynGO is expected to grow as an expert community effort**

Although SynGO 1.0 contains approximately 3000 annotations, this is still only a fraction of all relevant information available in scientific literature. Only for a core set of proteins, SynGO 1.0 contains three or more annotations per protein. A concerted effort by all experts involved in synapse research will help to uncover a larger fraction of available information on synapses and further improve the impact of SynGO. The publically accessible SynGO portal has been optimized to make such efforts with a user-friendly interface and stored credits for each annotator.

SynGO 1.0 contains 2922 annotations against 1112 genes, but proteomics studies of synaptic preparations implicate a few thousand proteins in synapses (Fig. 10). An unknown fraction of these synaptic candidate proteins will prove to be *bona fide* synaptic, for which the experimental evidence is currently lacking. It is important to note that biochemical preparations, based on the physical properties of proteins, cannot purify synapses or synaptic compartments to completeness and some candidate proteins will remain false positives. SynGO 1.0 does not include these candidates by default to avoid low confidence analyses with SynGO data. However, collections of candidate synaptic proteins inferred from proteomics studies, can be downloaded from the SynGO database and may guide the investigation of these candidates in future research by the synapse community. SynGO is also working together with Uniprot ([UniProt, 2018](#_ENREF_68)) to generate the necessary information on available antibodies to facilitate the validation of such candidates in the future. Furthermore, using the public SynGO interface (<https://syngoportal.org>), SynGO ontologies and gene annotations can be used for enrichment analyses of any new data set (genomic, mRNA or protein) and differences between experimental and control groups can be computed and visualized using SynGO visualization tools (Fig 1, Fig 2C-D).

Proteins that function in different types of synapses are systematically annotated in SynGO. However, SynGO 1.0 and currently published data do not yet provide sufficient resolution to define individual synaptic proteomes (synaptomes) down to specific synapse populations, which will be important to predict function, e.g. being facilitating or depressing, or being inhibitory or excitatory, and to identify alterations in disease. Biochemical purifications or other systematic studies of specific synapse populations will be required to establish such specific synaptomes. Until such data become available, the currently available single cell mRNA resources can be a proxy to define which synaptic genes are expressed in specific neuronal populations. Hence, continued research in the synapse field provides excellent opportunities to further improve and expand SynGO, while, conversely, SynGO can provide the conceptual framework and be a key hypothesis generator for such future studies.

**Conclusion**

Taken together, SynGO provides the scientific community with a public data resource for universal reference in synapse research, which is fully integrated in the Gene Ontology resource (<http://geneontology.org>), and thus ready for online gene enrichment analyses. SynGO annotations are included in the standard GO releases, and SynGO ontology terms are labeled in the GO ontology files as “goslim\_synapse”. By the engagement of the synapse research community, SynGO aims at reaching saturation to establish a truly comprehensive definition of the synapse. SynGO already brings together many expert laboratories, but actively seeks participation of additional experts to annotate new synaptic genes and/or refine existing annotations. A user-friendly interface (<https://syngoportal.org>) supports submission of such contributions, which will be reviewed by domain experts before being admitted to SynGO.

**STAR METHODS**

**Definition of (novel) synaptic gene ontologies and integration into Gene Ontology (GO)**

Ontology terms in SynGO v1.0 were compared to pre-existing synaptic ontologies in the GO database prior to the starting date of SynGO (2015-01-01). A snapshot of the GO database representing the state at 2015-01-01 was obtained from <http://purl.obolibrary.org/obo/go/releases/2014-12-22/go.obo> (the last release in 2014) and converted into a directed graph using the iGraph R package (<http://igraph.org>). To construct the CC and BP graphs in Fig 2 we first created a tree from the SynGO v1.0 ontologies and classified terms that were present in the GO snapshot as ‘reused’. Next, pre-existing synapse related terms that were not used by SynGO, indicated as purple nodes in Fig 2, were defined as child terms of these ‘reused’ terms within the GO snapshot. Finally, we restricted resulting terms to match the scope of SynGO v1.0 (typical glutamatergic and GABA-ergic synapses). Terms that further specialize parent terms into serotonic-, dopaminergic-, cholinergic-synapses, neuromuscular junctions, or ‘regulation of’ terms, were not taken into account in this evaluation of candidate terms for re-use by SynGO. Graphs in Fig 2 were visualized using a force-directed layout algorithm in Cytoscape([Shannon et al., 2003](#_ENREF_59)).

SynGO ontologies and annotations were integrated into the existing ontology structures within the GO database, and will continuously be updated as the SynGO project expands synaptic ontologies and adds annotations in the future. These GO ontologies are available in the ‘goslim\_synapse’ subset, its most recent version is always available at <http://purl.obolibrary.org/obo/go/subsets/goslim_synapse.obo>. Due to the nature of integrating domain specific ontologies into the GO database which has a much wider scope, some high-level SynGO ontology terms are not are literally adopted by GO. Respective SynGO annotations are translated when exported to GO, e.g., annotations against ‘process in the presynapse’ are stored in GO as ‘biological\_process(GO:0008150) occurs\_in presynapse(GO:0098793)’. The identifier of such terms that only exist in SynGO starts with “SYNGO:”, whereas terms also available in GO have identifiers that start with “GO:” (as seen in the SynGO terms list in Table S2). SynGO annotations as integrated into GO are available through existing GO tools and websites, analysis on the SynGO subset is possible by filtering for annotations with the ‘contributor=SynGO’ property. All data from the SynGO consortium together with purpose-built analysis tools and community engagement are available through the SynGO website at <https://www.syngoportal.org>.

**Gene expression data**

The “brain-expressed” control set consists of genes that were expressed in significantly higher levels in brain compared to other tissues in Genotype Tissue Expression Consortia (GTEx) data ([Ganna et al., 2016](#_ENREF_24)). The control set with “brain topN” was defined as the N highest expressed genes in brain, where N was set to the number of unique genes annotated in SynGO v1.0. The highest expressed genes were computed by ranking the average gene-expression levels (in RPKM) from all brain samples in GTEx ([Consortium et al., 2017](#_ENREF_16)) version 6 (*GTEx\_Analysis\_v6\_RNA-seq\_RNA-SeQCv1.1.8\_gene\_rpkm.gct.gz*).

For the brain enrichment analysis of synaptic genes in Fig 6 we computed the mean fold change comparing brain to all other tissues for each gene in the GTEx (version 7) data set. To examine enrichment, we applied a Wilcoxon Rank-Sum test for each SynGO ontology containing at least 5 genes. We used a one-sided hypothesis test in order to test whether the genes in the annotation are more brain expressed than expected under the null.

**Gene features**

Gene features described in Fig 4 and S4 were extracted from the BioMart ([Smedley et al., 2015](#_ENREF_60)) Ensembl Human genes GRCh38.p12 dataset and the Ensembl REST API Endpoints (release 95). Total gene length was computed using the start\_position and end\_position BioMart attributes (gene start and end, in base pairs). All known splice variants per gene were obtained through BioMart, from which the number of protein coding splice variants were counted using the transcript\_biotype attribute. cDNA length was extracted from gene sequences provided through the Ensembl REST API with ‘mask\_feature=1’ parameter, and analogously all transcript exonic and intronic regions were obtained.

**Isoform counts from full-length RNA sequencing**

From our recent publication ([Gupta et al., 2018](#_ENREF_29)) we isolated full-length long reads that were expressed in neuronal subtypes, namely external granular layer neurons, internal granular layer neurons and Purkinje cells and had been attributed to a spliced protein coding gene. Subsequently, we considered only genes that had 20 or more such reads and split this gene list into two subsets: those annotated in SynGO and its complement. These groups differed substantially in the number of reads per gene. In order to normalize this, we randomly selected 10 full-length reads for each gene, resulting in two gene lists (SynGO and non-SynGO) with exactly 10 reads each. We then counted the number of distinct isoforms that these 10 reads described for each gene and repeated this subsampling process 1000 times.

**Conservation of synaptic genes**

Cumulative distribution of genes by gene age: Gene trees, covering ~95% of human genes, were obtained from the PANTHER resource([Mi et al., 2018](#_ENREF_41)). Gene duplication events were dated relative to the earliest speciation node descending from the duplication. Trees were then pruned to contain only human paralogs, and the root of the tree (this ensures that fractional gene counts will add up to the total number of human genes). Each human gene was then traced back through the pruned tree to the root of the tree, and the number of branches was counted; this gives the total number of duplications (plus one, for the root) along the path to the root. Then, for each human gene, for each duplication (and root node) along the path from the gene to the root, a fractional count of 1/total was added to the count of genes that evolved at the date of that node. This process yields a count of human genes gained over each period of evolution, including gene duplication events. Estimated speciation times were taken from the TimeTree resource([Kumar et al., 2017](#_ENREF_35)). The tree of CPT1C-related genes was obtained from the PANTHER website and can be accessed, together with additional information about the sequences and a multiple sequence alignment, at <http://pantherdb.org/treeViewer/treeViewer.jsp?book=PTHR22589&species=agr>.

**Large scale protein-protein interaction data**

StringDB ([Szklarczyk et al., 2015](#_ENREF_65)) 10.5 human interactions (“9606.protein.links.detailed.v10.5.txt”) were filtered by combined score (700, high confidence) and experimental evidence (400, medium confidence). StringDB PPIs then were matched to SynGO HGNC annotated genes by gene symbol, or alternative names (“9606.protein.aliases.v10.5.txt”) for cases without a match. The distance between a pair of SynGO genes was defined as their path distance. For the CC model, the path distance between a membrane term and it’s integral, anchored or extrinsic sub-classes (e.g., from SV membrane to anchored component of SV membrane) was set to zero. For the null distribution we computed all path distances within the CC or BP graph between any pair of all SynGO genes.

**Proteomics of synaptic fractions**

Proteins identified in selected proteomics studies shown in Fig 10 were mapped to human gene identifiers (HGNC) using the <https://www.uniprot.org> ID mapping service and mapping tables provided through <https://www.genenames.org> (Table S4). Keratins were considered an external contaminant and therefore excluded from downstream analysis. The Venn diagram was generated using the ‘eulerr’ R package.

**GWAS datasets**

GWAS summary statistics for 8 traits were collected from the following resources; ADHD([Martin et al., 2018](#_ENREF_39)), Autism Spectrum Disorder, Bipolar Disorder([Bipolar et al., 2018](#_ENREF_9)) and Major Depressive Disorder([Wray et al., 2018](#_ENREF_71)) from <https://www.med.unc.edu/pgc/results-and-downloads>, Educational Attainment([Lee et al., 2018](#_ENREF_36)) from <https://www.thessgac.org/data>, Height([Wood et al., 2014](#_ENREF_70)) from <https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files>, Intelligence([Savage et al., 2018](#_ENREF_57)) from <https://ctg.cncr.nl/software/summary_statistics>, Schizophrenia([Pardinas et al., 2018](#_ENREF_47)) from <http://walters.psycm.cf.ac.uk/>.

**Magma geneset analysis**

First MAGMA gene analysis([de Leeuw et al., 2015](#_ENREF_18)) was performed using the 1000 Genome Phase3 reference panel for European population by assigning SNPs to genes within a 2kb upstream and 1kb downstream window for 20,319 genes. The default model (SNP-wide mean) was used. Then MAGMA gene-set analyses were then performed for SynGO and original synaptic GO terms. For SynGO, one additional set with all SynGO genes was added, and in total 154 terms with at least 5 annotated (unique) genes were tested. For original GO, 5 additional sets; all synaptic genes, all BP genes, all CC genes, presynapse and postsynapse were added, and in total 96 terms with at least 5 annotated (unique) genes were tested. The gene set analyses were performed with the following three conditions for each trait: 1) no additional covariate, 2) conditioning on brain and average expression across all tissue types based on GTEx v7 RNA-seq dataset ([Consortium et al., 2017](#_ENREF_16)), 3) conditioning on brain and average expression, and the level of conservation of the genes. GTEx v7 RNA-seq data was obtained from <https://gtexportal.org>. The homology conservation scores in Fig S7 represent the level of conservation of genes, measured by the number of species with homolog genes using 65 species available through BioMart. Bonferroni correction was performed for each analysis separately (*Pbon*=0.05/154 for SynGO and 0.05/96 for GO). Statistical results are available in Table S6.

**LDSC geneset analysis**

To assess the contribution of each SynGO term to disease/phenotype heritability, we applied Stratified LD-Score Regression (S-LDSC) (Finucane et al., 2015; Gazal et al., 2017) to binary gene set annotations constructed with a ±100KB window around each gene as done in previous work (Finucane et al., 2018; Zhu and Stephens, 2018). In our analyses, we conditioned on the 75 functional annotations in the baseline-LD model (Gazal et al., 2017), an annotation containing all 23,987 protein-coding genes with a ±100KB window, as well as brain-enriched genes (see above), and a continuous annotation representing the conservation score of each gene. For each gene set from SynGO or pre-existing synaptic GO annotations, we assessed the statistical significance of the gene set annotations standardized effect size 𝛕\*, (defined as the proportionate change in per-SNP heritability associated to a one standard deviation increase in the value of the annotation, conditioned on other annotations included in the model (Gazal et al., 2017)) based on Bonferroni correction. Statistical results are available in Table S6.

**ACKNOWLEDGMENTS:** SynGO was funded by The Stanley Center for Psychiatric Research at The Broad Institute of MIT and Harvard. SynGO was built on previous efforts (synaptic parts lists) funded by EU (EUROSPIN: HEALTH-F2-2009-241498 and SYNSYS: HEALTH-F4-2010-242167).

A.B.P was supported by Spanish grants ref. BFU2012-34398, BFU2015-69717-P (partially funded by FEDER funds of the EU), Ramón y Cajal Fellowship, ref. RYC-2011-08391, the European FP People: Marie Curie Action: Career Integration Grant, ref. 304111 and the CERCA Program/Generalitat de Catalunya. E.D.G and D.C.D were supported by DFG CRC779 project B09. M.V. was supported by an ERC Advanced grant from the EU (ERC-ADG-322966-DCVfusion).

**AUTHOR CONTRIBUTIONS:**

* Designed the study: G.F., S.E.H., F.K., P.v.N., A.B.S., P.D.T and M.V.
* Designed ontologies and reached consensus: all authors
* Implemented ontologies and evidence in GO, GO-training and quality control: R.F. B.K., R.L., H.M. and D.O-S.
* Annotated synaptic genes (>50): M.A-A, T.C., L.N.C., R.J.F., H.L.G., P.S.McPh, C.I., A.P.H. de J., H.J., M.K., N.L., H.MacG., P.v.N., M.N., V.O’C , R.P., K-H.S., R.F.G.T., C.V., R.R.V. and J.v.W.
* Supervised annotations: C.B., À.B., T.B., N.B., J.C., D.C.D., E.D.G.,C.H., R.L.H., R.J., P.S.K., E.K., M.R.K., P.S.McPh., V.O’C, T.A.R. and C.S.
* Performed annotation QC: F.K. and P.v.N.
* Performed bioinformatics analyses: A.B., D.P.H., F.K., H.T., K.T. and K.W.
* Supervised bioinformatics analyses: B.M.N., D.P., A.B.S, P.D.T. and M.V.
* Designed and built SynGO portal: F.K., with input of P.v.N., A.B.S and M.V.
* Generated figures: F.K. with input of A.B., D.P.H, P.v.N., K.T. and K.W.
* Wrote the paper: M.V., with input of T.C.B., F.K., A.B.S., P.D.T. and all expert laboratories

**DECLARATION of INTERESTS:**

The authors declare no competing interests. M.S. and C.H. were employees of Genentech, a member of the Roche Group.

**REFERENCES**

Abdou, K., Shehata, M., Choko, K., Nishizono, H., Matsuo, M., Muramatsu, S.I., and Inokuchi, K. (2018). Synapse-specific representation of the identity of overlapping memory engrams. Science *360*, 1227-1231.

Abul-Husn, N.S., Bushlin, I., Moron, J.A., Jenkins, S.L., Dolios, G., Wang, R., Iyengar, R., Ma'ayan, A., and Devi, L.A. (2009). Systems approach to explore components and interactions in the presynapse. Proteomics *9*, 3303-3315.

Arnsten, A.F., Wang, M.J., and Paspalas, C.D. (2012). Neuromodulation of thought: flexibilities and vulnerabilities in prefrontal cortical network synapses. Neuron *76*, 223-239.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T.*, et al.* (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature genetics *25*, 25-29.

Bayes, A., Collins, M.O., Croning, M.D., van de Lagemaat, L.N., Choudhary, J.S., and Grant, S.G. (2012). Comparative study of human and mouse postsynaptic proteomes finds high compositional conservation and abundance differences for key synaptic proteins. PLoS One *7*, e46683.

Bayes, A., Collins, M.O., Reig-Viader, R., Gou, G., Goulding, D., Izquierdo, A., Choudhary, J.S., Emes, R.D., and Grant, S.G. (2017). Evolution of complexity in the zebrafish synapse proteome. Nat Commun *8*, 14613.

Bayes, A., van de Lagemaat, L.N., Collins, M.O., Croning, M.D., Whittle, I.R., Choudhary, J.S., and Grant, S.G. (2011). Characterization of the proteome, diseases and evolution of the human postsynaptic density. Nature neuroscience *14*, 19-21.

Biesemann, C., Gronborg, M., Luquet, E., Wichert, S.P., Bernard, V., Bungers, S.R., Cooper, B., Varoqueaux, F., Li, L., Byrne, J.A.*, et al.* (2014). Proteomic screening of glutamatergic mouse brain synaptosomes isolated by fluorescence activated sorting. The EMBO journal *33*, 157-170.

Bipolar, D., Schizophrenia Working Group of the Psychiatric Genomics Consortium. Electronic address, d.r.v.e., Bipolar, D., and Schizophrenia Working Group of the Psychiatric Genomics, C. (2018). Genomic Dissection of Bipolar Disorder and Schizophrenia, Including 28 Subphenotypes. Cell *173*, 1705-1715 e1716.

Boda, B., Dubos, A., and Muller, D. (2010). Signaling mechanisms regulating synapse formation and function in mental retardation. Current opinion in neurobiology *20*, 519-527.

Bourgeron, T. (2015). From the genetic architecture to synaptic plasticity in autism spectrum disorder. Nature reviews Neuroscience *16*, 551-563.

Boyken, J., Gronborg, M., Riedel, D., Urlaub, H., Jahn, R., and Chua, J.J. (2013). Molecular profiling of synaptic vesicle docking sites reveals novel proteins but few differences between glutamatergic and GABAergic synapses. Neuron *78*, 285-297.

Calvo, S.E., Clauser, K.R., and Mootha, V.K. (2016). MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. Nucleic acids research *44*, D1251-1257.

Chang, R.Y., Etheridge, N., Nouwens, A.S., and Dodd, P.R. (2015). SWATH analysis of the synaptic proteome in Alzheimer's disease. Neurochem Int *87*, 1-12.

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.

Consortium, G.T., Laboratory, D.A., Coordinating Center -Analysis Working, G., Statistical Methods groups-Analysis Working, G., Enhancing, G.g., Fund, N.I.H.C., Nih/Nci, Nih/Nhgri, Nih/Nimh, Nih/Nida*, et al.* (2017). Genetic effects on gene expression across human tissues. Nature *550*, 204-213.

Consortium, U. (2018). UniProt: the universal protein knowledgebase. Nucleic acids research *46*, 2699.

de Leeuw, C.A., Mooij, J.M., Heskes, T., and Posthuma, D. (2015). MAGMA: generalized gene-set analysis of GWAS data. PLoS computational biology *11*, e1004219.

De Rubeis, S., He, X., Goldberg, A.P., Poultney, C.S., Samocha, K., Cicek, A.E., Kou, Y., Liu, L., Fromer, M., Walker, S.*, et al.* (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. Nature *515*, 209-215.

Demontis, D., Lescai, F., Borglum, A., Glerup, S., Ostergaard, S.D., Mors, O., Li, Q., Liang, J., Jiang, H., Li, Y.*, et al.* (2016). Whole-Exome Sequencing Reveals Increased Burden of Rare Functional and Disruptive Variants in Candidate Risk Genes in Individuals With Persistent Attention-Deficit/Hyperactivity Disorder. Journal of the American Academy of Child and Adolescent Psychiatry *55*, 521-523.

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. Nature neuroscience *11*, 799-806.

Filiou, M.D., Bisle, B., Reckow, S., Teplytska, L., Maccarrone, G., and Turck, C.W. (2010). Profiling of mouse synaptosome proteome and phosphoproteome by IEF. Electrophoresis *31*, 1294-1301.

Fromer, M., Pocklington, A.J., Kavanagh, D.H., Williams, H.J., Dwyer, S., Gormley, P., Georgieva, L., Rees, E., Palta, P., Ruderfer, D.M.*, et al.* (2014). De novo mutations in schizophrenia implicate synaptic networks. Nature *506*, 179-184.

Ganna, A., Genovese, G., Howrigan, D.P., Byrnes, A., Kurki, M., Zekavat, S.M., Whelan, C.W., Kals, M., Nivard, M.G., Bloemendal, A.*, et al.* (2016). Ultra-rare disruptive and damaging mutations influence educational attainment in the general population. Nature neuroscience *19*, 1563-1565.

Gaudet, P., Livstone, M.S., Lewis, S.E., and Thomas, P.D. (2011). Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium. Briefings in bioinformatics *12*, 449-462.

Giglio, M., Tauber, R., Nadendla, S., Munro, J., Olley, D., Ball, S., Mitraka, E., Schriml, L.M., Gaudet, P., Hobbs, E.T.*, et al.* (2018). ECO, the Evidence & Conclusion Ontology: community standard for evidence information. Nucleic acids research.

Grant, S.G. (2012). Synaptopathies: diseases of the synaptome. Current opinion in neurobiology *22*, 522-529.

Groschner, L.N., Chan Wah Hak, L., Bogacz, R., DasGupta, S., and Miesenbock, G. (2018). Dendritic Integration of Sensory Evidence in Perceptual Decision-Making. Cell *173*, 894-905 e813.

Gupta, I., Collier, P.G., Haase, B., Mahfouz, A., Joglekar, A., Floyd, T., Koopmans, F., Barres, B., Smit, A.B., Sloan, S.A.*, et al.* (2018). Single-cell isoform RNA sequencing characterizes isoforms in thousands of cerebellar cells. Nature biotechnology.

Heutink, P., and Verhage, M. (2012). Neurodegeneration: new road leads back to the synapse. Neuron *75*, 935-938.

Huang, K.Y., Su, M.G., Kao, H.J., Hsieh, Y.C., Jhong, J.H., Cheng, K.H., Huang, H.D., and Lee, T.Y. (2016). dbPTM 2016: 10-year anniversary of a resource for post-translational modification of proteins. Nucleic acids research *44*, D435-446.

Jeanquartier, F., Jean-Quartier, C., and Holzinger, A. (2015). Integrated web visualizations for protein-protein interaction databases. BMC bioinformatics *16*, 195.

Kandel, E.R. (2001). The molecular biology of memory storage: a dialogue between genes and synapses. Science *294*, 1030-1038.

Karczewski, K.J., Weisburd, B., Thomas, B., Solomonson, M., Ruderfer, D.M., Kavanagh, D., Hamamsy, T., Lek, M., Samocha, K.E., Cummings, B.B.*, et al.* (2017). The ExAC browser: displaying reference data information from over 60 000 exomes. Nucleic acids research *45*, D840-D845.

Kumar, S., Stecher, G., Suleski, M., and Hedges, S.B. (2017). TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. Molecular biology and evolution *34*, 1812-1819.

Lee, J.J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., Nguyen-Viet, T.A., Bowers, P., Sidorenko, J., Karlsson Linner, R.*, et al.* (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nature genetics *50*, 1112-1121.

Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B.*, et al.* (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285-291.

Lips, E.S., Cornelisse, L.N., Toonen, R.F., Min, J.L., Hultman, C.M., International Schizophrenia, C., Holmans, P.A., O'Donovan, M.C., Purcell, S.M., Smit, A.B.*, et al.* (2012). Functional gene group analysis identifies synaptic gene groups as risk factor for schizophrenia. Molecular psychiatry *17*, 996-1006.

Martin, J., Walters, R.K., Demontis, D., Mattheisen, M., Lee, S.H., Robinson, E., Brikell, I., Ghirardi, L., Larsson, H., Lichtenstein, P.*, et al.* (2018). A Genetic Investigation of Sex Bias in the Prevalence of Attention-Deficit/Hyperactivity Disorder. Biological psychiatry *83*, 1044-1053.

Mattheisen, M., Samuels, J.F., Wang, Y., Greenberg, B.D., Fyer, A.J., McCracken, J.T., Geller, D.A., Murphy, D.L., Knowles, J.A., Grados, M.A.*, et al.* (2015). Genome-wide association study in obsessive-compulsive disorder: results from the OCGAS. Molecular psychiatry *20*, 337-344.

Mi, H., Muruganujan, A., Ebert, D., Huang, X., and Thomas, P.D. (2018). PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic acids research.

Moczulska, K.E., Pichler, P., Schutzbier, M., Schleiffer, A., Rumpel, S., and Mechtler, K. (2014). Deep and precise quantification of the mouse synaptosomal proteome reveals substantial remodeling during postnatal maturation. J Proteome Res *13*, 4310-4324.

Monday, H.R., and Castillo, P.E. (2017). Closing the gap: long-term presynaptic plasticity in brain function and disease. Current opinion in neurobiology *45*, 106-112.

Morciano, M., Beckhaus, T., Karas, M., Zimmermann, H., and Volknandt, W. (2009). The proteome of the presynaptic active zone: from docked synaptic vesicles to adhesion molecules and maxi-channels. J Neurochem *108*, 662-675.

Morciano, M., Burre, J., Corvey, C., Karas, M., Zimmermann, H., and Volknandt, W. (2005). Immunoisolation of two synaptic vesicle pools from synaptosomes: a proteomics analysis. J Neurochem *95*, 1732-1745.

Pandya, N.J., Koopmans, F., Slotman, J.A., Paliukhovich, I., Houtsmuller, A.B., Smit, A.B., and Li, K.W. (2017). Correlation profiling of brain sub-cellular proteomes reveals co-assembly of synaptic proteins and subcellular distribution. Sci Rep *7*, 12107.

Pardinas, A.F., Holmans, P., Pocklington, A.J., Escott-Price, V., Ripke, S., Carrera, N., Legge, S.E., Bishop, S., Cameron, D., Hamshere, M.L.*, et al.* (2018). Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. Nature genetics *50*, 381-389.

Pedroso, I., Lourdusamy, A., Rietschel, M., Nothen, M.M., Cichon, S., McGuffin, P., Al-Chalabi, A., Barnes, M.R., and Breen, G. (2012). Common genetic variants and gene-expression changes associated with bipolar disorder are over-represented in brain signaling pathway genes. Biological psychiatry *72*, 311-317.

Petersen, C.C., and Crochet, S. (2013). Synaptic computation and sensory processing in neocortical layer 2/3. Neuron *78*, 28-48.

Phillips, G.R., Florens, L., Tanaka, H., Khaing, Z.Z., Fidler, L., Yates, J.R., 3rd, and Colman, D.R. (2005). Proteomic comparison of two fractions derived from the transsynaptic scaffold. J Neurosci Res *81*, 762-775.

Psychiatric, G.C.B.D.W.G. (2011). Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. Nature genetics *43*, 977-983.

Ripke, S., O'Dushlaine, C., Chambert, K., Moran, J.L., Kahler, A.K., Akterin, S., Bergen, S.E., Collins, A.L., Crowley, J.J., Fromer, M.*, et al.* (2013). Genome-wide association analysis identifies 13 new risk loci for schizophrenia. Nature genetics *45*, 1150-1159.

Ripolles, P., Ferreri, L., Mas-Herrero, E., Alicart, H., Gomez-Andres, A., Marco-Pallares, J., Antonijoan, R.M., Noesselt, T., Valle, M., Riba, J.*, et al.* (2018). Intrinsically regulated learning is modulated by synaptic dopamine signaling. eLife *7*.

Robinson, E.B., St Pourcain, B., Anttila, V., Kosmicki, J.A., Bulik-Sullivan, B., Grove, J., Maller, J., Samocha, K.E., Sanders, S.J., Ripke, S.*, et al.* (2016). Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. Nature genetics *48*, 552-555.

Roy, M., Sorokina, O., Skene, N., Simonnet, C., Mazzo, F., Zwart, R., Sher, E., Smith, C., Armstrong, J.D., and Grant, S.G.N. (2018). Proteomic analysis of postsynaptic proteins in regions of the human neocortex. Nature neuroscience *21*, 130-138.

Ruano, D., Abecasis, G.R., Glaser, B., Lips, E.S., Cornelisse, L.N., de Jong, A.P., Evans, D.M., Davey Smith, G., Timpson, N.J., Smit, A.B.*, et al.* (2010). Functional gene group analysis reveals a role of synaptic heterotrimeric G proteins in cognitive ability. American journal of human genetics *86*, 113-125.

Savage, J.E., Jansen, P.R., Stringer, S., Watanabe, K., Bryois, J., de Leeuw, C.A., Nagel, M., Awasthi, S., Barr, P.B., Coleman, J.R.I.*, et al.* (2018). Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. Nature genetics *50*, 912-919.

Selkoe, D.J. (2002). Alzheimer's disease is a synaptic failure. Science *298*, 789-791.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research *13*, 2498-2504.

Smedley, D., Haider, S., Durinck, S., Pandini, L., Provero, P., Allen, J., Arnaiz, O., Awedh, M.H., Baldock, R., Barbiera, G.*, et al.* (2015). The BioMart community portal: an innovative alternative to large, centralized data repositories. Nucleic acids research *43*, W589-598.

Smith, A.C., and Robinson, A.J. (2018). MitoMiner v4.0: an updated database of mitochondrial localization evidence, phenotypes and diseases. Nucleic acids research.

Soukup, S.F., Vanhauwaert, R., and Verstreken, P. (2018). Parkinson's disease: convergence on synaptic homeostasis. The EMBO journal *37*.

Spires-Jones, T.L., and Hyman, B.T. (2014). The intersection of amyloid beta and tau at synapses in Alzheimer's disease. Neuron *82*, 756-771.

Sudhof, T.C. (2008). Neuroligins and neurexins link synaptic function to cognitive disease. Nature *455*, 903-911.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P.*, et al.* (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic acids research *43*, D447-452.

Thapar, A., Martin, J., Mick, E., Arias Vasquez, A., Langley, K., Scherer, S.W., Schachar, R., Crosbie, J., Williams, N., Franke, B.*, et al.* (2016). Psychiatric gene discoveries shape evidence on ADHD's biology. Molecular psychiatry *21*, 1202-1207.

The Gene Ontology, C. (2018). The Gene Ontology Resource: 20 years and still GOing strong. Nucleic acids research.

UniProt, C. (2018). UniProt: a worldwide hub of protein knowledge. Nucleic acids research.

Wilhelm, B.G., Mandad, S., Truckenbrodt, S., Krohnert, K., Schafer, C., Rammner, B., Koo, S.J., Classen, G.A., Krauss, M., Haucke, V.*, et al.* (2014). Composition of isolated synaptic boutons reveals the amounts of vesicle trafficking proteins. Science *344*, 1023-1028.

Wood, A.R., Esko, T., Yang, J., Vedantam, S., Pers, T.H., Gustafsson, S., Chu, A.Y., Estrada, K., Luan, J., Kutalik, Z.*, et al.* (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. Nature genetics *46*, 1173-1186.

Wray, N.R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E.M., Abdellaoui, A., Adams, M.J., Agerbo, E., Air, T.M., Andlauer, T.M.F.*, et al.* (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nature genetics *50*, 668-681.

Yates, B., Braschi, B., Gray, K.A., Seal, R.L., Tweedie, S., and Bruford, E.A. (2017). Genenames.org: the HGNC and VGNC resources in 2017. Nucleic acids research *45*, D619-D625.

Zheng, J., Erzurumluoglu, A.M., Elsworth, B.L., Kemp, J.P., Howe, L., Haycock, P.C., Hemani, G., Tansey, K., Laurin, C., Early, G.*, et al.* (2017). LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. Bioinformatics *33*, 272-279.

Zwir, I., Arnedo, J., Del-Val, C., Pulkki-Raback, L., Konte, B., Yang, S.S., Romero-Zaliz, R., Hintsanen, M., Cloninger, K.M., Garcia, D.*, et al.* (2018). Uncovering the complex genetics of human temperament. Molecular psychiatry.

1. Dept. of Functional Genomics and [↑](#footnote-ref-1)
2. Dept. of Molecular and Cellular Neurobiology, Center for Neurogenomics and Cognitive Research, Vrije Universiteit Amsterdam, , 1081 HV Amsterdam, The Netherlands [↑](#footnote-ref-2)
3. Zilkha Neurogenetic Institute and Dept. of Psychiatry and Behavioral Sciences, Keck School of Medicine, University of Southern California, Los Angeles, CA 90333, USA [↑](#footnote-ref-3)
4. Institute of Cardiovascular Science, University College London, London WC1E 6JF, UK [↑](#footnote-ref-4)
5. European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Cambridge CB10 1SD, UK [↑](#footnote-ref-5)
6. Depts. of Neuroscience and Cell Biology, Howard Hughes Medical Institute, Kavli Institute for Neuroscience, Program in Cellular Neuroscience, Yale University School of Medicine, New Haven, CT 06520, USA [↑](#footnote-ref-6)
7. SIB Swiss Institute of Bioinformatics, Centre Medical Universitaire, Switzerland 1 rue Michel Servet, 1211 Geneva 4, Switzerland [↑](#footnote-ref-7)
8. Nancy Pritzker Laboratory, Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA 94305, USA [↑](#footnote-ref-8)
9. Dept. of Neuroscience, Genentech, South San Francisco, CA 94080, USA [↑](#footnote-ref-9)
10. Dept. of Molecular & Cellular Physiology, Howard Hughes Medical Institute, Stanford University, Stanford, CA 94305, USA [↑](#footnote-ref-10)
11. Brain and Mind Research Institute and Center for Neurogenetics, Weill Cornell Medicine, New York, New York, USA. [↑](#footnote-ref-11)
12. Dept. of Fundamental Neurosciences, University of Lausanne, CH-1006 Lausanne, Switzerland and Dept Biomedicine and Prevention, University of Rome Tor Vergata, 00133 Rome, Italy [↑](#footnote-ref-12)
13. Molecular Physiology of the Synapse Laboratory, Biomedical Research Institute Sant Pau, 08025 Barcelona, Spain and and Universitat Autònoma de Barcelona, 08193 Bellaterra, Cerdanyola del Vallès, Spain [↑](#footnote-ref-13)
14. Dept. of Neuroscience, Tufts University School of Medicine, Boston, MA 02111, USA [↑](#footnote-ref-14)
15. Dept. of Molecular Neurobiology, Max Planck Institute of Experimental Medicine, 37075 Göttingen, Germany [↑](#footnote-ref-15)
16. Dept. of Physiology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117597, Neurobiology/Ageing Program, Life Sciences Institute, National University of Singapore and Institute of Molecular and Cell Biology, Agency for Science, Technology and Research (A\*STAR), Singapore. [↑](#footnote-ref-16)
17. Leibniz Institute for Neurobiology (LIN); Center for Behavioral Brain Sciences (CBBS) and Medical Faculty, Otto von Guericke University, 39120 Magdeburg, Germany [↑](#footnote-ref-17)
18. Cell Biology, Dept. of Biology, Faculty of Science, Utrecht University, 3584 CH Utrecht, The Netherlands [↑](#footnote-ref-18)
19. Solomon H. Snyder Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA [↑](#footnote-ref-19)
20. Kavli Neuroscience Discovery Institute, Johns Hopkins University, Baltimore, MD 21205, US [↑](#footnote-ref-20)
21. Dept. of Neurobiology, Max-Planck-Institute for Biophysical Chemistry, 37077 Göttingen, Germany [↑](#footnote-ref-21)
22. Dept. of Neurobiology, Harvard Medical School, Boston, MA 02115, USA [↑](#footnote-ref-22)
23. Center for Synaptic Brain Dysfunctions, Institute for Basic Science (IBS) and Dept. of Biological Sciences, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, South Korea [↑](#footnote-ref-23)
24. RG Neuroplasticity, Leibniz Institute for Neurobiology, 39118 Magdeburg, Germany and Leibniz Group 'Dendritic Organelles and Synaptic Function', Center for Molecular Neurobiology, ZMNH, University MC, Hamburg-Eppendorf, Hamburg, 20251, Germany [↑](#footnote-ref-24)
25. Dept. of Neurology and Neurosurgery, Montreal Neurological Institute, McGill University, Montreal, Quebec H3A 2B4, Canada [↑](#footnote-ref-25)
26. Analytic and Translational Genetics Unit, Dept. of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114; Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA [↑](#footnote-ref-26)
27. Biological Sciences, University of Southampton, Southampton, SO17 1BJ, UK [↑](#footnote-ref-27)
28. Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Vrije Universiteit Amsterdam, and Dept. of Clinical Genetics, Amsterdam UMC, 1081 HV Amsterdam, The Netherlands [↑](#footnote-ref-28)
29. Dept. of Biochemistry, Weill Cornell Medicine, New York, NY 10065, USA [↑](#footnote-ref-29)
30. CNR Neuroscience Institute Milan and Dept. of Biotechnology and Translational Medicine, University of Milan, 20129 Milan, Italy [↑](#footnote-ref-30)
31. McGovern Institute for Brain Research, Dept. of Brain and Cognitive Sciences, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA [↑](#footnote-ref-31)
32. Harvard Dept. of Stem Cell and Regenerative Biology, Stanley Center for Psychiatric Research, Cambridge, MA 02138, USA [↑](#footnote-ref-32)
33. Division of Bioinformatics, Dept. of Preventive Medicine, Keck School of Medicine of USC, University of Southern California, Los Angeles, CA 90033, USA [↑](#footnote-ref-33)