**Supplemental information**

**Contents**

1. TAP tagged Arc vector and gene targeting.
2. TAP tagged Arc transgenic mice.
3. TAP tagged Arc distribution and subcellular localization.
4. Electrophysiological parameters measured in the CA1 area of ArcTAP/TAP hippocampal slices.
5. Classification of Arc interactors according to their gene ontology process.
6. Arc protein levels in the forebrain of PSD93/Dlg2 and SAP102/Dlg3 knockout mice.
7. Arc binding to PSD95/Dlg4 protein is independent of PSD93/Dlg2.
8. Isolation of TAP tagged Arc complexes from PSD95/Dlg4 mutant mice.
9. Methods:

9.1 Tandem Affinity Purification

9.2 Proteomic analysis.

9.3 Network building.

9.4 Genetic analysis.

* De novo mutation exome sequence datasets
* Enrichment analysis of rare coding mutations in Arc interactors in human neuropsychiatric disease.
* Creating human Arc interactor genesets.

9.5 Human cognitive ability phenotype

9.6 Cellular fractionation.

9.7 Statistics.

10. Supplementary Table legends

11. References.

1. **TAP tagged Arc vector and gene targeting.**

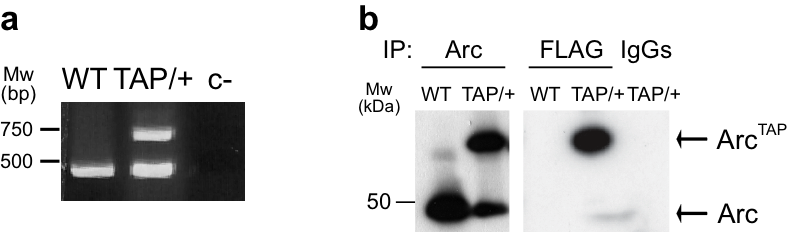
The TAP tag fragment containing the Histidine Affinity Tag (HAT), TEV protease and Flag sequences flanked by XbaI and BclI restriction sites was cloned into pneoflox vector (TAPtag pneoflox) as described in [1](#_ENREF_1). Two homology arms of the genomic Arc/Arg3.1 sequence were amplified with forward ArcHAXhoIF and reverse ArcHAXbaIR primers and forward ArcHAAcc65IF and reverse ArcHABglIIR primers using the BAC bMQ101B9 as a template. Both homology arms were cloned into the TAPtag pneoflox vector leaving in between the TAP tag sequence, 2 loxP sites, PGK and EM7 promoters, the G418r gene and an SV40 polyadenylation site. The cassette flanked by two homology arms was removed and transformed into *EL350* *E. coli* cells containing a pTargeter vector with the genomic Arc/Arg3.1 sequence cr1574496136 to cr15 74506490 (Ensembl release 49). The cassette was inserted into the pTargeter vector by recombination[2](#_ENREF_2).The final vector contained a 5’-end homolog Arc sequence of 6664bp and a 3’-end homology arm of 3688bp(ENSMUSG00000022602). The vector was linearized with PvuI and electroporated into E14 embryonic stem cells. The neomycin resistant colonies were picked up, expanded and frozen. Genomic DNA was extracted from all of them and PCR with and pneoF3 and Arc4R (Supplementary Table 12) to identify TAP Arc homologous recombinants.



**Supplementary Figure 1.** Scheme of the genomic Arc locus targeted with the TAP tag (Figure 1). The TAP sequence was inserted before the stop codon of the protein. The cross with a transgenic Cre-expressing mouse line allowed to deleting the neomycin resistance cassette (neo) by recombination between *loxP*-sites (bottom). Asterisk: stop codon (TAA) of the coding sequence; Black thick lane: TAP tag sequence; Triangle: *loxP* site.

**2. TAP tagged Arc transgenic mice.**

One of the ES cells positive clones, FA7, was microinjected into C57BL/6 blastocysts and generated 4 germline chimaeras containing 30-90% of targeted cells. These chimaeras were crossed onto a C57BL/6 genetic background. Tail DNA from the litters was extracted and analyzed by PCR with a 5’ Arc1F primer and two 3’ pneo3R and Arc10R primers to distinguish the heterozygous (+/-) and wild type (+/+) alleles, respectively. The knockin strain is referred as ArcTAP. ArcTAP heterozygous mice (ArcTAP/+) did not show any distortion of transmission frequency in the offspring of Arc intercrosses. TAP tagged Arc protein levels were similar to wild type Arc levels in forebrain extracts of ArcTAP/+ mice and TAP tagged Arc was specifically immunoprecipitated by specific anti-Arc and anti-FLAG antibodies. Mice used for experimental analysis were 2- to 5-month-old males from several intercrosses between the chimeras and C57BL/6 strain mice. All animal experiments were conducted in a licenced animal facility in accordance to guidelines determined through the UK Animals (Scientific Procedures) Act, 1986 and all procedures were approved through the British Home Office Inspectorate. Animals were born and housed at the Wellcome Trust Sanger institute, and exposed to similar light/dark cycles and food supply. Mice were sacrificed through cervical dislocation, and the forebrain was dissected on ice and snap frozen in liquid nitrogen. Brain samples were stored at -80ºC for few weeks prior to use.



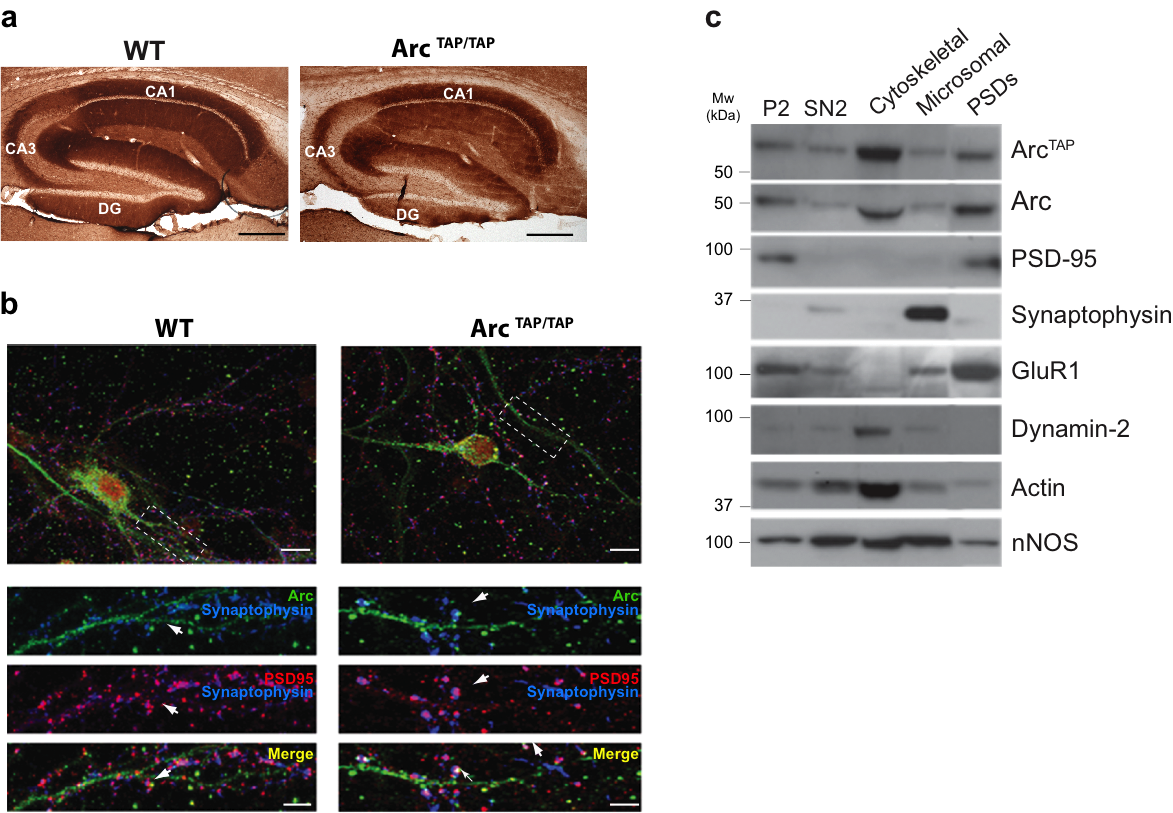
**Supplementary Figure 2.** (a) PCR that amplifies the wild type (lower band) and targeted alleles (upper band). (b) TAP tagged Arc was specifically purified from ArcTAP/+ forebrain extracts with an anti-Arc and anti-Flag antibodies and blotted with an Arc antibody. WT, wild type; TAP/+, heterozygous for TAP tagged Arc; c-, PCR water; IgG, mouse total IgG used as a negative control of the immunoprecipitation.

**3. TAP tagged Arc distribution and subcellular localization.**

TAP tagged Arc showed the same distribution pattern in dendrites of CA1, CA3 and Dentate Gyrus of hippocampus as wild type Arc[3](#_ENREF_3) (Supplementary Figure 3A) Mice were anesthetized with a mix of ketamine/xylacine and cardiac perfused with 0.1 M sodium phosphate buffer (pH 7.4) followed by 4 % Paraformaldehyde in 0.1M sodium phosphate buffer (pH 7.4). The brain was removed and placed in fixative for a further 1 h and then equilibrated overnight in 30 % sucrose in PBS. Free-floating sections were incubated with anti-Arc (1:250; Santa Cruz Biotechnologies). Sections were incubated with biotinylated anti-mouse secondary antibody (1:200; DAKO) for peroxidase staining. The signal was amplified using an avidin-biotin system (Vector Laboratories) followed by a staining in 3,3-diaminobenzidine (Sigma).

TAP tagged Arc showed a similar subcellular distribution as wild type Arc as it showed partial colocalization with PSD95 and synaptophysin in synaptic boutons. Hippocampal primary neurons were dissected from mouse embryos (embryonic stage 17.5 days) wild type and ArcTAP/TAP C57Bl/6 littermates and digested in papain (Supplementary Figure 3B). The culture conditions and the immnocytochemistry protocol were performed as described in [4](#_ENREF_4).Primary neurons were fixed and stained at 14 days in vitro (DIV) and labeled with the following antibodies:rabbit Arc 1/250 (Santa Cruz Biotechnologies), IgG2A mouse PSD-95 1/350 (Affinity Bioreagents) and IgG1 mouse Synaptophysin 1/1000 (EMD Milipore). Secondary antibodies conjugated with Alexa Fluor dyes (Invitrogen) were applied against the specific IgGs. All images were taken on a Zeiss 510 META confocal microscope using a 63x Plan-apochromat objective. Z-stacks were taken at 0.8 m intervals and maximal projections were made to give the images showed.

Biochemical fractionation of ArcTAP/TAP and wild type forebrains showed that TAP tagged Arc was enriched in PSD and cytoskeletal fractions as was wild type Arc (Supplementary Figure 3C). Brain fractionation was done according to the protocol described in [5](#_ENREF_5). Briefly, mouse forebrains were homogenized in 10 volumes of sucrose buffer (0.32 M sucrose, 4 mM HEPES, pH 7.4, 1 mM EDTA, 1 mM EGTA and protease inhibitors cocktail tablets (Roche)) with a glass Teflon Dounce homogenizer, and centrifuged at 800g for 15 min. The supernatant (SN1) was centrifuged 15 min at 9000g and the pellet collected as crude membrane fraction (P2). The supernatant (SN2) containing the cytosolic fraction was again centrifuged 60 min at 100.000g. The pellet was collected as microsomal fraction (P4) and the supernatant (SN4) was considered as the cytoskeletal fraction. P2 crude membrane fraction containing synaptosomes was solubilized in 0.2% DOC buffer and protease inhibitors and centrifuged 20 min at 100.000g. Pellet (P3) was considered as PSD-enriched fraction.



**Supplementary Figure 3.** (a) Hippocampal sections of wild type and ArcTAP/TAP mice stained with anti-Arc antibody.. *DG*: Dentate gyrus. Bar=1 mm. (b) Representative image of embryonic primary neurons derived from wild type and ArcTAP/TAP mice immunostained at DIV15 with antibodies against Arc (green), PSD-95 (red) and Synaptophysin (blue). Merged: colocalization of the three signals. Arrows show the puncta labeling with each protein. Bar=1 m.(c)Biochemical fractionation from ArcTAP/TAP and wild type forebrain mice. Similar protein amount from each fraction was loaded onto a gel and immunoblotted with the antibodies displayed on the right of the panel. TAP tagged Arc showed the same subcellular distribution as the wt non-tagged isoform. Antibodies against GluR1, Synaptophysin, Dynamin-3, Actin and nNOS were used as specific markers of each fraction. Mw: molecular weight in kDa.

**4. Electrophysiological parameters measured in the CA1 area of ArcTAP/TAP hippocampal slices.**

Previous studies show that disruption of Arc/Arg3.1 led to disruption of synaptic long-term potentiation (LTP) and long-term depression (LTD) in the hippocampal CA1 area and dentate gyrus [6-8](#_ENREF_6). In addition, Arc/Arg3.1 is required for normal basal synaptic transmission, as knock-out of the *Arc* gene or its over-expression in neuronal cultures alter synaptic strength possibly by regulating endocytosis and trafficking of AMPA receptor subunits[5](#_ENREF_5),[6](#_ENREF_6),[9](#_ENREF_9),[10](#_ENREF_10). In view of these data, we sought to explore if incorporation of the TAP tag into Arc led to disturbances of synaptic transmission and plasticity. Using previously described multi-electrode array technology[11](#_ENREF_11), we compared input-output relationship, paired-pulse facilitation and theta-burst induced LTP in CA1 area of acute hippocampal slices prepared from ArcTAP/TAP mice and their wild type littermates. Acute hippocampal slices were used to record field excitatory post synaptic potentials (fEPSPs), by the MEA60 electrophysiological suite (Multi Channel Systems, Reutlingen, FRG) [11](#_ENREF_11),[12](#_ENREF_12). Eight set-ups consisting of a MEA1060-BC pre-amplifier and a filter amplifier (gain 550x) were run simultaneously by data acquisition units operated by MC\_Rack software. Raw electrode data were digitized at 10 kHz and stored on a PC hard disk for subsequent analysis. To record fEPSPs, a hippocampal slice was placed into the well of 5x13 3D multi electrode array (MEA) biochip (Qwane Biosciences, Lausanne, Switzerland). The slice was guided to a desired position with a fine paint brush and gently fixed over MEA electrodes by a silver ring with attached nylon mesh lowered vertically by a one-dimensional U-1C micromanipulator (You Ltd, Tokyo, Japan). MEA biochips were fitted into the pre-amplifier case and fresh ACSF was delivered to the MEA well through a temperature-controlled perfusion cannula that warmed perfused media to 32°C. Monopolar stimulation of Schäffer collateral/commissural fibers through array electrodes was performed by STG4008 stimulus generator (Multi Channel Systems, Reutlingen, FRG). Biphasic (positive/negative, 100 µs/a phase) voltage pulses were used. Amplitude, duration and frequency of stimulation were controlled by MC\_Stimulus II software. All experiments were performed using two-pathway stimulation of Schäffer collateral/commissural fibers. Our previous experiments that utilized MEAs, demonstrated that largest LTP was recorded in proximal part of apical dendrites of CA1 pyramidal neurons [11](#_ENREF_11). We have therefore picked a single principal recording electrode in the middle of proximal part of CA1 and assigned two electrodes for stimulation of the control and test pathways on the subicular side and on the CA3 side of *stratum radiatum* respectively. The distance from the recording electrode to the test stimulation electrode was 400-510 µm and to the control stimulation electrode 316-447 µm. To evoke orthodromic fEPSPs, test and control pathways were activated in succession at a frequency of 0.02 Hz. Baseline stimulation strength was adjusted to evoke a response that corresponded to 40% of the maximal attainable fEPSP at the recording electrode located in proximal *stratum radiatum*. Amplitude of the negative part of fEPSPs was used as a measure of the synaptic strength. Paired stimulation with an interpulse interval of 50 ms was used to observe paired-pulse facilitation (PPF) in baseline conditions in the test pathway before LTP induction. PPF was calculated by dividing the amplitude of the fEPSP obtained in response to the second pulse by the amplitude of fEPSP evoked by the preceding pulse. To induce LTP, 10 bursts of baseline strength stimuli were administered at 5 Hz to test pathway with 4 pulses given at 100 Hz per burst (total 40 stimuli). LTP plots were scaled to the average of the first five baseline points. Normalisation of LTP values was performed by dividing the fEPSP amplitude in the tetanised pathway by the amplitude of the control fEPSP at corresponding time points. Normalised LTP values averaged across the period of 61-65 min after theta-burst stimulation were used for statistical comparison.None of the three measured parameters was statistically different between the two groups of mice. Therefore, we conclude that insertion of TAP tag into Arc/Arg3.1 does not lead to major alterations of its synaptic function.



**Supplementary Figure 4:** (A) Basal synaptic transmission was normal in ArcTAP/TAP mice. Input-output relationships (A1) illustrate averaged fEPSP amplitudes in slices from ArcTAP/TAP (*nslice*s = 15, *Nanimals* = 5) and wild type mice (*n* = 18, *N* = 5) in response to stimulation of Schäffer collaterals. AUCI-O values were not statistically different in ArcTAP/TAP and mutant animals (F(1,7.06) = 0.258; *P* = 0.627). Representative families of fEPSP traces are given in A2. (B) Paired-pulse facilitation (B1) was not statistically different (F(1,7.36) = 2.405; *P* = 0.163) in ArcTAP/TAP animals (*n* = 15, *N* = 5) as compared to their wild type littermates (*n* = 18, *N* = 5). Representative fEPSP sweeps are given in B2. (C) Theta-burst stimulation elicited pathway specific long-term potentiation of synaptic transmission in hippocampal CA1 area (C1). Normalised magnitude of this potentiation 60-65 min after LTP induction did not differ in mutant mice (166 ± 5%; *n* = 15, *N* = 5; F(1,7.75) = 0.449; *P* = 0.522) relatively to their wild type counterparts (171±4%; *n* = 18, *N* = 5). Examples of test pathway fEPSP traces immediately before and 60 min after theta-burst stimulation are presented in C2. WT= wild type.

**5. Classification of Arc interactors according to their gene ontology process.**

Proteins in the Arc complexes (107 proteins, 102 from single step and 39 tandem purifications) were classified according to their Biological Process (Supplementary Tables 5 and 6) using the QuickGO browser[13](#_ENREF_13) and the database InterPro2Go (Mike Croning, database?)

****

**Supplementary Figure 5.** Bar diagram with the number of proteins purified in the single step and the tandem purifications according to their gene ontology process.

**6. Arc protein levels in the forebrain of PSD93/Dlg2 and SAP102/Dlg3 knockout mice.**

Forebrain extracts were solubilized in 1% DOC buffer at 0.38 g wet weight per 7 ml cold buffer with a glass Teflon Douncer homogenizer. The homogenate was incubated 1h at 4ºC and clarified at 50000 *g* for 30 min at 4ºC. Protein concentration was quantified with the Bradford protein-assay (BioRad). Samples containing equal amount of protein were subjected to reducing SDS electrophoresis (NUPAGE, Invitrogen) and transferred to polyvinyldifluoride membrane (HybondTM-P, GE Healthcare). The membranes were blocked in 5% non-fat milk, 0.01% Triton X-100 in PBS with the antibodies: mouse Arc (Santa Cruz Biotechnology), mouse SAP102 (NeuroMab), mouse PSD93 (NeuroMab), mouse SAP97 (BD Biosciences), rabbit GluR1 (EMD Millipore), and nNos (EMD Millipore) Detection of signals was carried out using peroxidase-linked secondary IgGs (Jackson) and enhanced chemiluminescence (GE Healthcare). Quantification of the signals was performed using ImageJ open source program.

**Supplementary Figure 6.** Arc protein levels in forebrain of PSD-93/Dlg2 and SA102/Dlg3 mutant mice. Representative immunoblot showing relative abundance of different proteins in lysates of hippocampus from PSD-93/Dlg2 and SA102/Dlg3 knockout mice and matched wild type littermates. Western blots were probed with antibodies to PSD-95, SAP102, PSD-93, Arc, GluR1, GluR2 and nNOS as indicated. Quantification of protein expression was calculated to the amount of nNOS in each sample and expressed as a percentage of the wild type expression. -/-: knockout. N= 4. \* *p*< 0.05, Mann-Whitney *U* test.



**7. Arc binding to PSD95/Dlg4 protein is independent of PSD93/Dlg2**

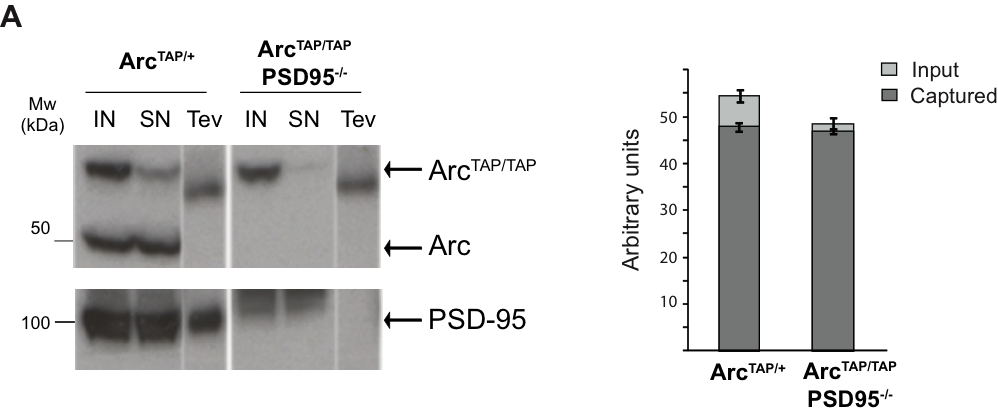
Based on the LC-MS/MS data on the tandem purification of Arc complexes (Supplementary Table 1), PSD95/Dlg4 and PSD93/Dlg2 showed high abundance within Arc complexes whereas no peptide was found for SAP102/Dlg3. We verified the interaction of Arc with PSD95 and checked whether thjs interaction was affected by absence of PSD93/Dlg2 immunoprecipitation. Mouse forebrains from PSD95/Dlg4 knockout and PSD93/Dlg2 knockout and their wild type littermates were homogenized in 1% DOC buffer following the same protocol as in the TAP procedure. Equal amount of protein extracts were incubated for 2h at 4ºC with mouse PSD95 (Affinity Bioreagents) Captured complexes were then washed three times with DOC buffer and eluted with 1x NuPAGE LDS sample buffer (Invitrogen) supplemented with reducing agent. Absence of PSD93/Dlg2 did not affect the interaction of Arc to PSD95 suggesting that PSD-95-Arc interaction is independent of PSD93.

**Supplementary Figure 7:** Arc interaction to PSD95 is independent of PSD93. PSD95 was immunoprecipitated from PSD95-/- and PSD93-/- mutant and wild type mice forebrain extracts and immunoblotted against Arc and PSD95. IP: antibodies used for immunoprecipitation, hc: Antibody heavy chain. -/-: knockout mice.



**8. Isolation of Arc TAP/TAP complexes from PSD95/Dlg4 mutant mice.**

ArcTAP/TAP mice were crossed with PSD-95 knockout mouse [14](#_ENREF_14) in order to generate the knockin/knockout strain ArcTAP/TAPxPSD95-/-. Arc binding partners were captured from lysed forebrain from ArcTAP/TAPxPSD-95-/-. In order to isolate similar amount of Arc-associated protein complexes from wild type and PSD95-/-, TAP tagged Arc was captured from heterozygous ArcTAP/+ mice. The TAP procedure, concentration of the eluted samples and gel staining were performed as described in Experimental procedures. The amount of protein lysed, captured and eluted was monitored by semiquantitative immunoblotting before LC-MS/MS.

****

**Supplementary Figure 8**. Arc complexes were isolated from ArcTAP/+ and ArcTAP/TAP crossed with PSD95knockout mice (ArcTAP/TAPxPSD95-/-) by Flag capture and Tev protease release (single step purification). (**A**) Total lysate (IN, input), same volume of lysate upon the purification (SN, supernatant) and Tev elution from both genotypes were blotted against Arc and quantified. Eluted Arc levels following the Flag capture from ArcTAP/+ and ArcTAP/TAPxPSD-95-/- lysateswere not statistically different, *p* =0.1). (B) Isolated complexes from (A) were resolved by SDS-PAGE and stained with colloidal Coomassie. Three independent purifications are shown. The lanes were cut for LC-MS/MS analysis and the identified proteins listed in Supplementary Table 7. TAP tagged Arc, PSD95 and the Tev enzyme are indicated.

**9. Methods**

**9.1 Tandem Affinity Purification**

Mouse forebrain was homogenized on ice in 1% DOC buffer (50 mM Tris pH9.0, 1% Sodium Deoxycholate, 50 mM NaF, 20M ZnCl, 1mM Na3VO4, 2 mM Pefabloc SC (Roche) and 1 tablet/10ml protease inhibitor cocktail tablets (Roche) at 0.38 g wet weight per 7 ml cold buffer with a glass Teflon Douncer homogenizer. The homogenate was incubated 1h at 4ºC and clarified at 50000 *g* for 30 min at 4ºC. Isolation of Arc TAP tagged complexes was performed as described [1](#_ENREF_1). Briefly, clarified protein extracts were incubated with Flag antibody coupled to Dynal beads (Invitrogen) for 2 h at 4ºC. Captured complexes were then washed with DOC buffer and twice with TEV-protease cleavage buffer (Invitrogen) supplemented with 0.1% Deoxycholate. Arc TAP tagged protein was cut off by addition of TEV protease (Invitrogen) at 4ºC overnight. For the tandem purification, the eluate was incubated with NI-NTA agarose beads (Quiagen) at 4ºC for 40 min. The beads were then washed three times with washing buffer (50 mM Sodium Phosphate pH 8.0, 50 mM NaCl, 0.1% sodium deoxycholate and 1mM of Imidazole). The protein complexes were eluted with 250 mM Immidazole. To assess the efficiency of the TEV elution and the tandem purification, all the beads were boiled in SDS sample buffer after protein elution. The TEV eluate as well as the imidazole eluted fractions were concentrated in a Vivaspin concentrator (Vivascience, GE), reduced with DTT, alkylated with iodoacetamide and separated by one-dimensional SDS-electrophoresis 4-12% (NUPAGE, Invitrogen, CA). The gel was fixed, stained with Colloidal Coomassie and lanes were cut into slices, destained and digested overnight with trypsin (Roche, Trypsin modified, sequencing grade) as in [1](#_ENREF_1).

**9.2 Proteomic analysis**

LC-MS/MS analysis

The TEV eluate as well as the imidazole eluted fractions were concentrated in a Vivaspin concentrator (Vivascience, GE), reduced with DTT, alkylated with iodoacetamide and separated by one-dimensional SDS-electrophoresis 4-12% (NUPAGE, Invitrogen, CA). Protein gels were stained overnight with colloidal Coomassie blue (Sigma). Each lane was excised into 12 bands that were destained and in-gel digested overnight using trypsin (sequencing grade; Roche). Peptides were extracted from gel bands twice with 50% acetonitrile/0.5% formic acid and dried in a SpeedVac (Thermo). Peptides were resuspended using 0.5% formic acid were analysed using an Ultimate 3000 Nano/Capillary LC System (Dionex) coupled to either an LTQ FT Ultra (Characterisation of TAP-Arc complexes) or an LTQ Orbitrap Velos (differential analysis of TAP-ARC complexes in wild type or PSD-95 mutants) hybrid mass spectrometers (Thermo Electron) equipped with a nanospray ion source. Peptides were desalted on-line using a micro-Precolumn cartridge (C18 Pepmap 100, LC Packings) and then separated using either a 35 min RP gradient (4-32% acetonitrile/0.1% formic acid) on a BEH C18 analytical column (1.7um, 75 μm id x 10 cm,) (Waters) (Characterisation of TAP-Arc complexes) or a 125 min RP gradient (4-32% acetonitrile/0.1% formic acid) on an Acclaim PepMap100 C18 analytical column (3 μm, 75 μm id x 50 cm,) (Dionex) (differential analysis of TAP-ARC complexes in wild type or PSD-95 mutants). The mass spectrometer was operated in standard data dependent acquisition mode controlled by Xcalibur 2.0/2.1. The LTQ-FT Ultra was operated with a cycle of one MS (in the FTICR cell) acquired at a resolution of 100,000 at m/z 400, with the top five most abundant multiply-charged (2+ and higher) ions in a given chromatographic window subjected to MS/MS fragmentation in the linear ion trap. The LTQ-Orbitrap Velos was operated with a cycle of one MS (in the Orbitrap) acquired at a resolution of 60,000 at m/z 400, with the top 20 most abundant multiply-charged (2+ and higher) ions in a given chromatographic window subjected to MS/MS fragmentation in the linear ion trap. FTMS target values of 5e5 (LTQ-FT Ultra) and 1e6 (LTQ-Orbitrap Velos) and an ion trap MSn target values of 1e4 were used. Maximum FTMS scan accumulation times were set at 1000ms (LTQ-FT Ultra) and 500ms (LTQ-Orbitrap Velos) and maximum ion trap MSn scan accumulation tiles were set to 200ms (LTQ-FT Ultra) and 100ms (LTQ-Orbitrap Velos). Dynamic exclusion was enabled with a repeat duration of 45s with an exclusion list of 500 and exclusion duration of 30s.

Data analysis

Data was analysed data using MaxQuant version 1.3.0.5[15](#_ENREF_15). MaxQuant processed data was searched against a UniProt mouse proteome sequence database (Mar, 2012) using following search parameters: trypsin with a maximum of 2 missed cleavages, 7 ppm for MS mass tolerance, 0.5 Da for MS/MS mass tolerance, with Acetyl (Protein N-term) and Oxidation (M) set as variable modifications and carbamidomethyl (C) as a fixed modification. A protein FDR of 0.01 and a peptide FDR of 0.01 were used for identification level cut offs. Protein quantification was performed using razor and unique peptides and using only unmodified and carbamidomethylated peptides.

Summary of one-step purifications

Two sets of one-step purifications:

3 replicates of ARCTAP/TAP and wild type (as control) purifications dimethyl labelled for quantification by MaxQuant).

2 replicates of ARCTAP/TAP and 1 control (MaxQuant intensity-based label free quantification)

Proteins were considered enriched if they had a ratio of enrichment of 2.5 (ARCTAP/TAP/Control purification) in two out of 5 one-step purifications.

Summary of tandem purifications

3 replicates of ARCTAP/TAP and control purifications (MaxQuant intensity-based label free quantification). Proteins were considered enriched if they had a ratio of enrichment of 2.5 (ARCTAP/TAP/Control purification) in two out of three two-step purifications. The total set of 107 ARC complex components was assembled from the combined set of enriched proteins from one step and tandem purifications.

Summary of purification from ARCTAP/TAP/PSD95-/- mice

3 replicate one-step purifications of ARC complexes from ARCTAP/+, ARCTAP/TAP/PSD95-/- and control mice were dimethyl labelled for quantification by MaxQuant. Given that ARC expression is down-regulated in PSD-95 mutant mice, all protein ratios were normalized to that determined for ARC by MaxQuant. 77 out of the total set of 107 ARC complex proteins were quantified in dimethyl labelled PSD95 mutant versus wild type experiments and 59 of these proteins could be quantified in at least 2 out of 3 replicates. Of these, 24 were decreased by 1.5-fold in the PSD-95 mutant (Ratio < 0.67) and 11 showed a 1.5 fold increase in abundance in the PSD95 mutant (Ratio >1.5).

**9.3 Network building**

**9.4 Human genetic analysis**

***De novo* mutation exome sequencing datasets.**

We obtained publicly available exome sequencing *de novo* mutation data from a variety of sources, yielding a total of 3,985 autism trios[16-18](#_ENREF_16), 356 epilepsy trios[4](#_ENREF_4), 192 ID trios[19-21](#_ENREF_19) and 1024 schizophrenia trios[22-26](#_ENREF_22). As negative controls, we also obtained similar data from 2,049 healthy controls[17](#_ENREF_17),[20](#_ENREF_20),[25](#_ENREF_25),[26](#_ENREF_26) and 362 congenital heart disease trios[27](#_ENREF_27).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Disease** | ***N* trios** | ***N* LoF mutations** | ***N* nonsynonymous mutations** | **Sources (PMIDs)** |
| Autism | 3,985 | 579 | 3446 | 25363760, 23849776, 25363768 |
| Epilepsy | 356 | 58 | 341 | 25262651 |
| ID | 192 | 67 | 259 | 23033978, 25356899, 23020937 |
| Schizophrenia | 1024 | 114 | 756 | 24463507, 21743468, 23911319, 24776741, 23042115 |

The Swedish schizophrenia case/control association data were based on summary statistics from a previously described dataset[28](#_ENREF_28). We used PLINK/Seq (<http://atgu.mgh.harvard.edu/plinkseq/>) to uniformly re-annotate all mutations, with respect to RefSeq genes. Consistent with the previously reported work, for all *de novo* studies we focused on two classes of mutation: 1) gene disruptive (or loss-of-function, LoF) mutations (i.e. nonsense, essential splice site or frameshift indels); 2) a broad class of all nonsynonymous mutations (including disruptive mutations). For the case/control schizophrenia study, we focused on 1) rare gene disruptive mutations with a sample frequency of less than 0.1%; 2) rare (<0.1%) nonsynonymous mutations that were predicted *in silico* to have a likely deleterious effect on protein function, following[28](#_ENREF_28).

**Enrichment analysis of rare coding mutations in Arc interactors in human neuropsychiatric disease.**

We used the same analytic strategies (implemented in freely-available software) as previously described. For the analysis of *de novo* mutations, we used DNENRICH[22](#_ENREF_22)(<https://psychgen.u.hpc.mssm.edu/dnenrich/>). Briefly, DNENRICH estimates the observed number of *de novo* mutations per gene or geneset by a genomic permutation strategy, which controls for gene size and structure, sequence coverage and local trinucleotide mutation rate. Significance of the enrichment in the rate and recurrence of mutations was assessed empirically by permutation. For the case/control data, we used the SMP algorithm[28](#_ENREF_28) which is part of the PLINK/Seq package (<http://atgu.mgh.harvard.edu/plinkseq/>). Briefly, the test for enrichment of a set of genes is based on the sum of gene-level case/control burden statistics relative to the exome-wide excess of burden in cases. Significance is assessed by permutation, comparing the observed distribution against 10,000 null replicates (created by shuffling case/control labels within groups matched for ancestry and technical variables).

**Creating human Arc interactor genesets.**

From the original 107 Arc interactor mouse proteins, we identified homologous human genes based on the Mouse Genome Informatics database (<http://www.informatics.jax.org/faq/ORTH_dload.shtml>). After mapping and manual curation of gene identifiers, we found 124 human Arc interactors RefSeq genes, of which 13 were associated with increased Arc in the PSD95-/- mouse, 37 of which were associated with decreased Arc, and 18 of which were known direct interactors of PSD95 (Supplementary Table 7).

***De novo* CNV enrichment**

In order to map rodent proteins onto human genes, all of the majority protein ids from Table X were first converted into both MGI & mouse NCBI/Entrez gene ids using the online ID mapping tool provided by Uniprot. These gene ids were then converted to human Entrez ids using the mapping file 'HOM\_MouseHumanSequence.rpt', available from MGI (<http://www.informatics.jax.org/>). Any genes with a non-unique (e.g. 1-many) mapping between species, or where MGI and mouse Entrez ids mapped to different human genes, were excluded.

For each gene set, the number of genes hit by 34 *de novo* CNVs from schizophrenia case were compared those hit by 59 *de novo* CNVs from unaffected Icelandic controls. A gene was counted as being hit by a CNV if the CNV overlapped any part of its length (human genome Build 36.3). To overcome biases related to gene and CNV size, and to control for differences between studies and genotyping chips, the following logistic regression models were fitted to the combined set of CNVs:

(a) logit (pr(case)) = CNV size + total number of genes hit

(b) logit (pr(case)) = CNV size + total number of genes hit + number of genes hit in gene set

Comparing the change in deviance between models (a) and (b), a one-sided test for an excess of genes in the gene set being hit by case CNVs was performed. The genetic data used for this analysis is described in [29](#_ENREF_29). Note that while a small number of the protein groups from Table X correspond to multiple proteins/genes, their presence does not influence the analysis as none were hit by either case or control CNVs.

**9.5 Human cognitive ability phenotype**

The data processing pipeline used for the construction of the general cognitive ability phenotype in the five discovery cohorts and the replication cohorts has been detailed previously[30](#_ENREF_30). In brief, genome wide association was carried out in each cohort of CAGES using Mach2QTL[31](#_ENREF_31) before being meta-analysed in METAL [32](#_ENREF_32" \o "Willer, 2010 #68) using an inverse variance weighted model. SNPs were then assigned to genes based on their position in the UCSC human genome browser hg 18 assembly with a 50kb boundary around each gene to capture any regulatory elements. A gene based statistic was derived using VEGAS [33](#_ENREF_33) to control for the number of SNPs assigned to each gene as well as patterns of linkage disequilibrium. In order to test the principal hypothesis that the genes of the Arc complex will show a greater association, as a set, than those drawn from across the genome, each gene based p value was –log10 transformed and rank ordered. Gene Set Enrichment Analysis (GSEA) [34](#_ENREF_34),[35](#_ENREF_35), a competitive test of enrichment, was then used to determine if the gene identifiers for the Arc gene set fall higher in the genome-wide rankings than would be expected by chance alone [30](#_ENREF_30),[36](#_ENREF_36). This was done for each set by deriving a Kolmogorov-Smirnov (K-S) statistic weighted by the p-values of the gene based statistic in order to take into account both the ranks and the distance between ranks. Following this the genome wide ranked set was permuted and the K-S statistic calculated again. Statistical significance was established by using 15,000 permutations of the genome wide ranked set with the p-value describing the proportion of permuted K-S tests smaller than the original un-permuted K-S statistic. Statistical significance was set at < 0.05 and FDR <0.25 [34](#_ENREF_34),[35](#_ENREF_35).

**9.6 Cellular fractionation**

Brain cellular fractionation was done according to the protocol described in [5](#_ENREF_5). Briefly, mouse forebrains were homogenized in 10 volumes of sucrose buffer (0.32 M sucrose, 4 mM HEPES, pH 7.4, 1 mM EDTA, 1 mM EGTA and protease inhibitors cocktail tablets (Roche)) with a glass Teflon Dounce homogenizer, and centrifuged at 800g for 15 min. The supernatant (SN1) was centrifuged 15 min at 9000g and the pellet collected as crude membrane fraction (P2). The supernatant (SN2) containing the cytosolic fraction was again centrifuged 60 min at 100.000g. The pellet was collected as microsomal fraction (P4) and the supernatant (SN4) was considered as the cytoskeletal fraction. P2 crude membrane fraction containing synaptosomes was solubilized in 0.2% DOC buffer and protease inhibitors and centrifuged 20 min at 100.000g. Pellet (P3) was considered as PSD-enriched fraction.

Arc TAP tagged complexes were isolated from the cytoskeletal fraction (SN4) following a Tandem Affinity Purification. Briefly, two ArcTAP/TAP mouse forebrains were pulled and homogenized in sucrose buffer. Following a cytoskeletal fractionation, SN4 was incubated with Flag antibody coupled to Dynal beads (Invitrogen) at 4ºC for 2 h. The TAP procedure was performed as above.

**9.7 Statistics**

Hippocampal slices

In electrophysiological experiments, area under the input-output relationship curve (AUCI-O) was calculated to assess the effect of the Arc/Arg3.1TAP mutation on the basal synaptic transmission. Since several slices were routinely prepared from every mouse, AUCI-O, PPF and LTP values were compared between wild-type and mutant mice using the two-way nested ANOVA with genotype (group) and mice (sub-group) as fixed and random factors correspondingly (STATISTICA v. 10, StatSoft, Inc., Tulsa, OK, USA). DF error was computed using the Satterthwaite method and main genotype effect was considered significant if *p* < 0.05. Graph plots and normalisation were performed using OriginPro 8.5 (OriginLab, Northampton, MA, USA). Electrophysiological data are presented as mean ± s.e.m. with *n* and *N* indicating number of slices and mice respectively.

Protein quantification

Comparisons between two groups were performed using two-tailed Mann-Whitney *U* test. Significance was accepted to *p* < 0.05. Data are presented as mean ± s.e.m.

**10. Supplementary Tables legends**

**Supplementary Table 1.** Proteins identified in a single step and TAP purifications. Protein IDs (UniProt accession numbers), protein and gene names, approved peptide numbers, peptide spectrum matches (PSMs) and intensities for each protein identified by LC-MS/MS are shown. Protein intensity ratios (purification over control) as well as number of replicates are indicated. Arc interactors common with previous PSD95 purifications (columns BD, BE) and known PSD95 interactors (column BF) are also indicated. See Supplementary Methods for peptide and protein approval criteria.

1S: Single step purification (Columns G, H, J, K, L); Cont: Control purifications from wild type mice (columns I, P Q, R); Tandem: Tandem Affinity purification (columns M, N, O).

**Supplementary Table 2. Known Arc interactors.** Arc interactors previously reported have been manually annotated. Protein IDs (UniProt accession numbers), and protein and gene names are shown. PubMed ID and the interaction detection method are also indicated. Interactors fount in Arc complexes are highlighted.

**Supplementary Table 3. Intensity-based absolute quantification (iBAQ) numbers of Arc and PSD95/Dlg4.** iBAQ numbers of Arc and PSD95 from two independent single step purifications (ARC14, ARC15) and three independent Tandem purifications (ARC02, ARC03, ARC05) are shown. 1S: single step purification.

**Supplementary Table 4.** **Intensity-based absolute quantification (iBAQ) numbers of Arc-core interactors.** iBAQ numbers of Arc interactors in three independent Tandem purifications (ARC02, ARC03, ARC05) are shown.

**Supplementary Table 5. Classification of 107 Arc interactors according to their protein class.** Protein IDs (UniProt accession numbers) and protein and gene names are indicated. Protein were classified according to its biological process and molecular function ontologies using the QuickGo browser[13](#_ENREF_13) and the database InterPro2Go. Biological process ontologies (also indicated in Supplementary Table 6) were arranged to the following protein classes: Adaptor/Regulatory, Chaperone/Protein folding/ Signaling, Cytoskeletal/Structural/Cell adhesion, DNA/RNA binding, Enzyme, G-protein signaling, Ion channel, Phosphatase, Receptor/ Transporter, Ser/Thr Kinase, Vesicular/Trafficking/Transport and Unclassified. Number of purification replicates for each protein is indicated. 1S: Single step purification; Tandem: Tandem Affinity purification.

**Supplementary Table 6. Biological process ontology of 107 Arc interactors.** Biological processes of Arc interactors are shown. Enrichment of each process is also indicated. Synaptic transmission, ion and cation transport, and regulation of synaptic transmission are among the most statistical significant biological processes.

**Supplementary Table 7.** Proteins identified in single step purifications from ArcTAP/+ and ArcTAP/TAP crossed with PSD95/Dlg4 knockout mice. Protein IDs (UniProt accession numbers), protein and gene names, approved peptide numbers, diME ratios and intensity ratios for each protein identified by LC-MS/MS are shown. Number of Arc purification replicates (Columns V,Z), PSD95 purification replicates (Column AG), and known PSD95 interactors (column AH) are also indicated. Dimethyl labeling ratios are indicated as L (Control purification), M (ArcTAP/+ purification) and H (ArcTAP/TAPPSD95-/- purifications). Purification replicates are named as PSD95mut01, PSD95mut02 and PSD95mut03. The averaged ratio H/M has been normalized by the amount of Arc in the ArcTAP/TAP PSD95-/- purifications (Column X). See supplementary methods for peptide and protein approval criteria.

1S: Single step purification; Cont: Control purifications from wild type mice; Tandem: Tandem Affinity purifications. -/-: knockout mouse.

Mark, what’s the difference between columns G,H,I and J,K,L?

**Supplementary Table 8. Arc gene set enrichment in schizophrenia, epilepsy, autism and intellectual disability.** Summary of Arc genes interrogated for loss of function (LoF) and nonsynonymous mutations. Arc interaction type refers to its increment a reduction in PSD95/Dlg4 knockout mouse. AUT: Autism; EPI: Epilepsy; ID: Intellectual disability; SCZ: Schizophrenia, N; non; Y: Yes.

**Supplementary Table 9.** **Map of Arc complex to OMIM disorders.** Arc complex gene names, OMIM IDs and diseases are indicated.

**Supplementary Table 10.** **Mammalian Phenotype Ontology of Arc interactors**. Enrichment of phenotypes are indicated. Abnormal synaptic transmission, nervous system physiology and spatial learning are among the most statistical significant phenotypes.

**Supplementary Table 11. Arc gene set association with general cognitive ability.** Arc gene sets (107 interactors) and human chromosome location including start and stop nucleotides are shown not including the 50 kb boundary included around each gene. Number of SNPs for each gene annotated in the Brisbane Adolescent Twin Study (BATS) and Cognitive Ageing in England and Scotland consortium are also indicated.

**11. References**

1 Fernandez, E. *et al.* Targeted tandem affinity purification of PSD-95 recovers core postsynaptic complexes and schizophrenia susceptibility proteins. *Molecular systems biology* **5**, 269, doi:10.1038/msb.2009.27 (2009).

2 Liu, P., Jenkins, N. A. & Copeland, N. G. A highly efficient recombineering-based method for generating conditional knockout mutations. *Genome Res* **13**, 476-484 (2003).

3 Steward, O., Wallace, C. S., Lyford, G. L. & Worley, P. F. Synaptic activation causes the mRNA for the IEG Arc to localize selectively near activated postsynaptic sites on dendrites. *Neuron* **21**, 741-751 (1998).

4 Valor, L. M., Charlesworth, P., Humphreys, L., Anderson, C. N. & Grant, S. G. Network activity-independent coordinated gene expression program for synapse assembly. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 4658-4663, doi:10.1073/pnas.0609071104 (2007).

5 Chowdhury, S. *et al.* Arc/Arg3.1 interacts with the endocytic machinery to regulate AMPA receptor trafficking. *Neuron* **52**, 445-459, doi:10.1016/j.neuron.2006.08.033 (2006).

6 Shepherd, J. D. *et al.* Arc/Arg3.1 mediates homeostatic synaptic scaling of AMPA receptors. *Neuron* **52**, 475-484, doi:10.1016/j.neuron.2006.08.034 (2006).

7 Guzowski, J. F. *et al.* Inhibition of activity-dependent arc protein expression in the rat hippocampus impairs the maintenance of long-term potentiation and the consolidation of long-term memory. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **20**, 3993-4001 (2000).

8 Messaoudi, E. *et al.* Sustained Arc/Arg3.1 synthesis controls long-term potentiation consolidation through regulation of local actin polymerization in the dentate gyrus in vivo. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **27**, 10445-10455, doi:10.1523/JNEUROSCI.2883-07.2007 (2007).

9 Rial Verde, E. M., Lee-Osbourne, J., Worley, P. F., Malinow, R. & Cline, H. T. Increased expression of the immediate-early gene arc/arg3.1 reduces AMPA receptor-mediated synaptic transmission. *Neuron* **52**, 461-474, doi:S0896-6273(06)00738-0 [pii]

10.1016/j.neuron.2006.09.031 (2006).

10 Waung, M. W., Pfeiffer, B. E., Nosyreva, E. D., Ronesi, J. A. & Huber, K. M. Rapid translation of Arc/Arg3.1 selectively mediates mGluR-dependent LTD through persistent increases in AMPAR endocytosis rate. *Neuron* **59**, 84-97 (2008).

11 Kopanitsa, M. V., Afinowi, N. O. & Grant, S. G. Recording long-term potentiation of synaptic transmission by three-dimensional multi-electrode arrays. *BMC neuroscience* **7**, 61, doi:10.1186/1471-2202-7-61 (2006).

12 Coba, M. P. *et al.* TNiK is required for postsynaptic and nuclear signaling pathways and cognitive function. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **32**, 13987-13999, doi:10.1523/JNEUROSCI.2433-12.2012 (2012).

13 Binns, D. *et al.* QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics* **25**, 3045-3046, doi:10.1093/bioinformatics/btp536 (2009).

14 Migaud, M. *et al.* Enhanced long-term potentiation and impaired learning in mice with mutant postsynaptic density-95 protein. *Nature* **396**, 433-439, doi:10.1038/24790 (1998).

15 Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology* **26**, 1367-1372, doi:10.1038/nbt.1511 (2008).

16 De Rubeis, S. *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209-215, doi:10.1038/nature13772 (2014).

17 Iossifov, I. *et al.* De novo gene disruptions in children on the autistic spectrum. *Neuron* **74**, 285-299, doi:S0896-6273(12)00340-6 [pii]

10.1016/j.neuron.2012.04.009 (2012).

18 Jiang, Y. H. *et al.* Detection of clinically relevant genetic variants in autism spectrum disorder by whole-genome sequencing. *American journal of human genetics* **93**, 249-263, doi:10.1016/j.ajhg.2013.06.012 (2013).

19 de Ligt, J. *et al.* Diagnostic exome sequencing in persons with severe intellectual disability. *The New England journal of medicine* **367**, 1921-1929, doi:10.1056/NEJMoa1206524 (2012).

20 Rauch, A. *et al.* Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet* **380**, 1674-1682, doi:10.1016/S0140-6736(12)61480-9 (2012).

21 Hamdan, F. F. *et al.* De novo mutations in moderate or severe intellectual disability. *PLoS genetics* **10**, e1004772, doi:10.1371/journal.pgen.1004772 (2014).

22 Fromer, M. *et al.* De novo mutations in schizophrenia implicate synaptic networks. *Nature* **506**, 179-184, doi:10.1038/nature12929 (2014).

23 McCarthy, S. E. *et al.* De novo mutations in schizophrenia implicate chromatin remodeling and support a genetic overlap with autism and intellectual disability. *Molecular psychiatry* **19**, 652-658, doi:10.1038/mp.2014.29 (2014).

24 Girard, S. L. *et al.* Increased exonic de novo mutation rate in individuals with schizophrenia. *Nature genetics* **43**, 860-863, doi:10.1038/ng.886 (2011).

25 Gulsuner, S. *et al.* Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. *Cell* **154**, 518-529, doi:10.1016/j.cell.2013.06.049 (2013).

26 Xu, B. *et al.* De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. *Nature genetics* **44**, 1365-1369, doi:10.1038/ng.2446 (2012).

27 Zaidi, S. *et al.* De novo mutations in histone-modifying genes in congenital heart disease. *Nature* **498**, 220-223, doi:10.1038/nature12141 (2013).

28 Purcell, S. M. *et al.* A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* **506**, 185-190, doi:10.1038/nature12975 (2014).

29 Kirov, G. *et al.* De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Molecular psychiatry* **17**, 142-153, doi:10.1038/mp.2011.154 (2012).

30 Hill, W. D. *et al.* Human cognitive ability is influenced by genetic variation in components of postsynaptic signalling complexes assembled by NMDA receptors and MAGUK proteins. *Transl. Psychiatry* **4**, e341 (2014).

31 Li, Y., Willer, C. J., Ding, J., Scheet, P. & Abecasis, G. R. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genetic epidemiology* **34**, 816-834, doi:10.1002/gepi.20533 (2010).

32 Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-2191, doi:10.1093/bioinformatics/btq340 (2010).

33 Liu, J. Z. *et al.* A versatile gene-based test for genome-wide association studies. *American journal of human genetics* **87**, 139-145, doi:10.1016/j.ajhg.2010.06.009 (2010).

34 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U S A* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).

35 Wang, K., Li, M. & Bucan, M. Pathway-based approaches for analysis of genomewide association studies. *American journal of human genetics* **81**, 1278-1283, doi:10.1086/522374 (2007).

36 Hill, W. D. *et al.* Functional Gene Group Analysis Indicates No Role for Heterotrimeric G Proteins in Cognitive Ability. *PloS one* **9** (2014).