SUPPLEMENTARY MATERIAL

Disruption of the *ASTN2* / *TRIM32* locus at 9q33.1 is a risk factor in males for Autism Spectrum Disorders, ADHD and other neurodevelopmental phenotypes.

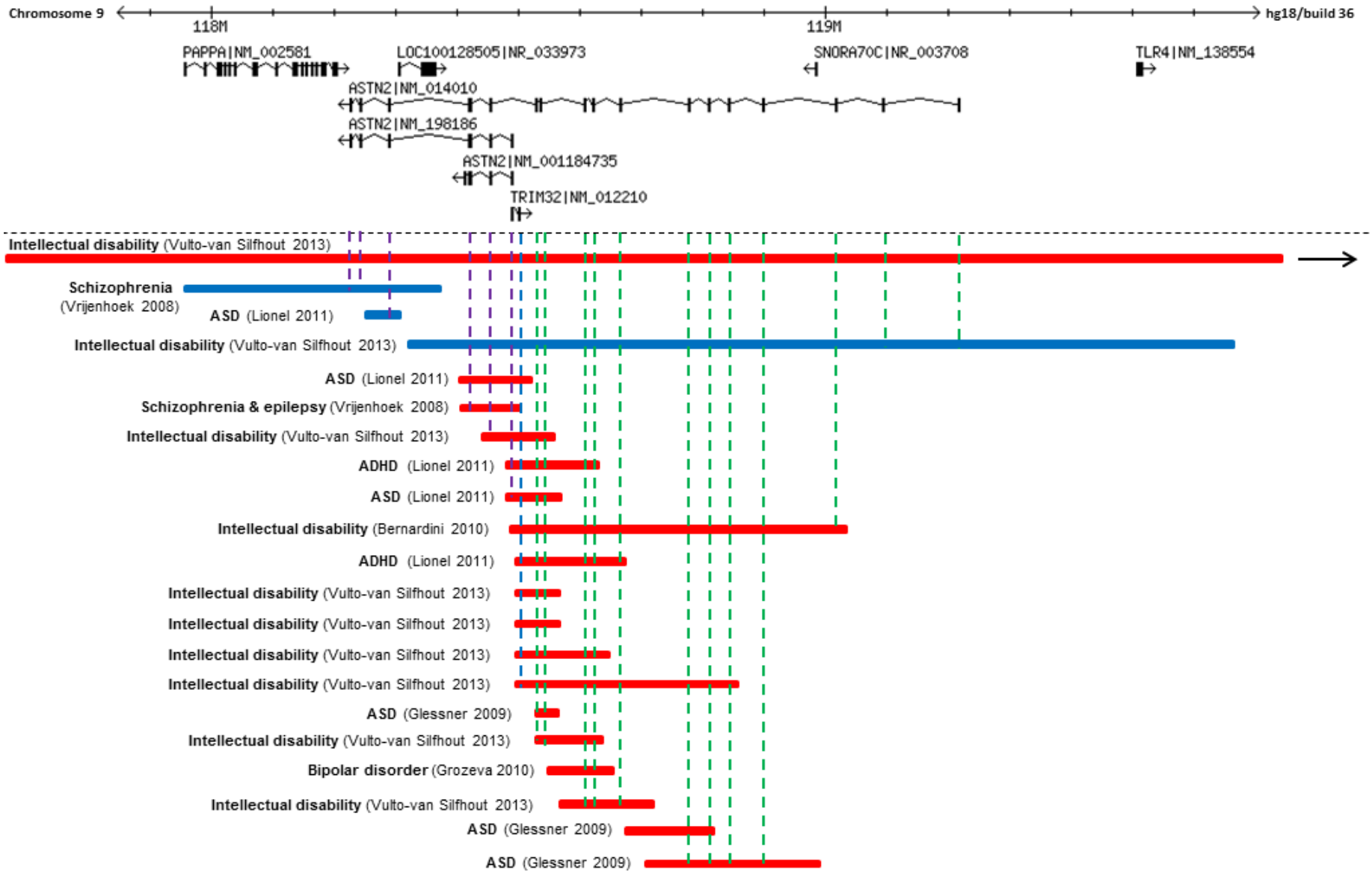
Chromosome 9
118M
119M
hg18/build 36
PAPPA|NM_002581
LOC100128505|NR_033973
SNORA70C|NR_003708
TLR4|NM_138554
ASTN2|NM_014010
ASTN2|NM_198186
ASTN2|NM_001184735
TRIM32|NM_012210
Intellectual disability (Vulto-van Silfhout 2013)
Schizophrenia (Vrijenhoek 2008)
ASD (Lionel 2011)
Intellectual disability (Vulto-van Silfhout 2013)
ASD (Lionel 2011)
Schizophrenia & epilepsy (Vrijenhoek 2008)
Intellectual disability (Vulto-van Silfhout 2013)
ADHD (Lionel 2011)
ASD (Lionel 2011)
Intellectual disability (Bernardini 2010)
ADHD (Lionel 2011)
Intellectual disability (Vulto-van Silfhout 2013)
Intellectual disability (Vulto-van Silfhout 2013)
Intellectual disability (Vulto-van Silfhout 2013)
Intellectual disability (Vulto-van Silfhout 2013)
ASD (Glessner 2009)
Intellectual disability (Vulto-van Silfhout 2013)
Bipolar disorder (Grozeva 2010)
Intellectual disability (Vulto-van Silfhout 2013)
ASD (Glessner 2009)
ASD (Glessner 2009)

Figure S1. Exonic CNVs reported at the *ASTN2/TRIM32* locus by previous studies.

Blue and red bars represent duplications and deletions respectively. Dashed purple lines intersect CNVs that overlap exons shared by multiple *ASTN2* isoforms and green lines intersect those affecting only the long isoform. Dashed vertical blue line intersects CNVs that overlap an exon of *TRIM32*. Genomic locations and coordinates are based on hg18 (NCBI36). Information about genes and transcript isoforms was obtained from the RefSeq database. The three transcript isoforms of *ASTN2* possessing different number of exons are depicted including the long isoform (NM_014010) and two shorter isoforms (NM_198186 and NM_001184735). The three other shorter isoforms of the gene (NM_198187, NM_198188 and NM_001184734) have the same number and location of exons as NM_198186 but differ slightly in the length of their first and terminal exons and UTRs. CNVs presented in this figure are from the following genome-wide microarray scans of different neurodevelopmental disorders: Vrijenhoek *et al*.,(1) Lionel *et al*.,(2) Bernardini *et al*.,(3) Glessner *et al*.(4), Grozeva *et al*.(5) and Vulto-van Silfhout *et al*.(6) Coordinates were not specified by Fernandez *et al*.(7) for the exonic deletions reported in two individuals with Tourette syndrome. The five exonic CNVs from Lionel et al. are included, together with more clinical information, in this study as patients 14, 21, 22, 26 and 48. The significant enrichments of exonic deletions affecting multiple *ASTN2* isoforms in NDD cases vs, controls ($p = 0.009$) and in male NDD cases vs. male controls ($p = 0.015$) are still observed if these cases are removed from the CNV analyses.

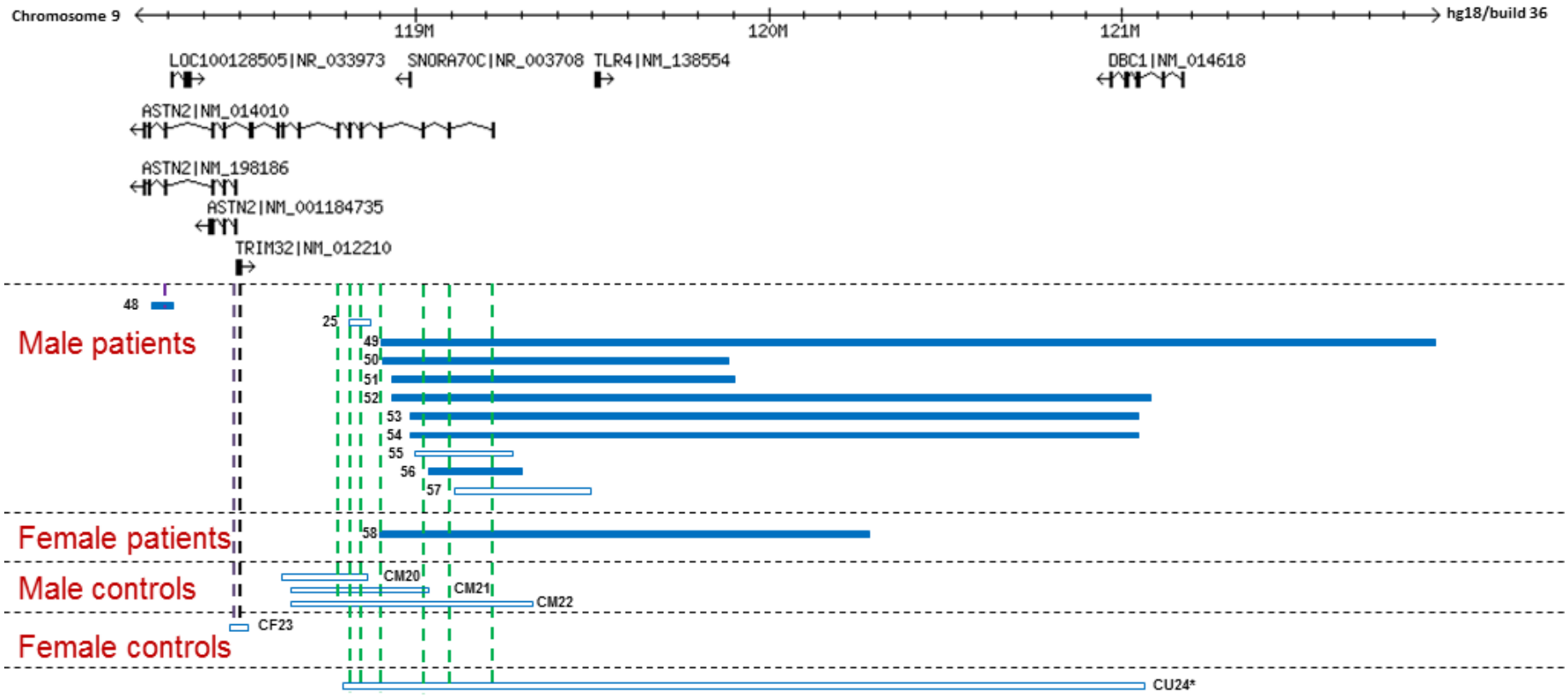

Figure S2. Exonic gains at the *ASTN2/TRIM32* locus in clinical and control cohorts.

Exonic duplications identified in 12 / 89,985 cases and 5 / 44,085 controls are depicted. All described duplication CNVs in the cases, except for the one in male patient 48, involve only the 5' portion of the longest isoform of *ASTN2*. These duplications were not found to be significantly enriched in cases vs. controls and their etiological contribution is unclear. Filled blue bars represent duplications detected in individuals with NDD phenotypes. Empty blue bars denote duplications in cases without known NDD traits from available clinical information and in controls. Numbers adjacent to the bars are the randomized sample ids of individuals with the duplications and correlate with information in Table 2, Table S1 and Table S3. Gender information was not available for the control individual (CU24) marked with * at the bottom of the figure. Dashed purple lines intersect duplications that overlap exons shared by multiple *ASTN2* isoforms and green lines intersect duplications affecting only the long isoform. Dashed vertical black line intersects duplications that overlap an exon of *TRIM32*. Genomic locations and coordinates are based on hg18 (NCBI36). Information about genes and transcript isoforms was obtained from the RefSeq database.

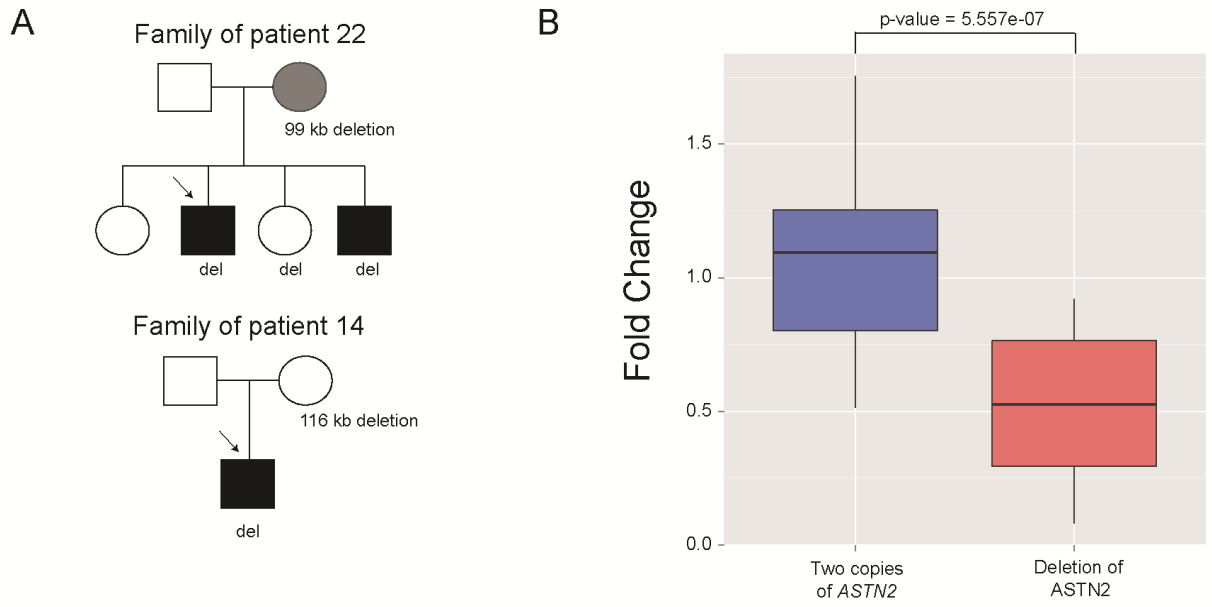

Figure S3. Functional impact of *ASTN2* deletions on gene expression in lymphoblasts.

A) Pedigrees of the two families used for *ASTN2* expression analysis. Arrows indicate the index probands of the pedigrees (denoted as patients 14 and 22 in this study). Individuals with black shading have an ASD diagnosis while the person represented by gray shading reported depression and anxiety issues. B) Relative *ASTN2* expression in lymphoblast cell lines from the six individuals (3 males and 3 females) with *ASTN2* deletions in the two pedigrees were compared to lymphoblast cell-lines from 9 individuals (5 males and 4 females) with two copies of *ASTN2*. The latter group included the three individuals without deletions in the two pedigrees as well as six additional samples with two copies of *ASTN2*. The lymphoblast cell lines were cultured as previously described.(8) Total RNA was extracted using Qiagen RNeasy mini kit with DNase I treatment (QIAGEN, Valencia, CA, USA). cDNA synthesis and qRT-PCR were performed as described in the main Methods. The expression was measured using two different primer pairs which detected all *ASTN2* isoforms (Table S10). *ASTN2* expression was normalized using *GAPDH* (dCt) and the fold change was calculated using the ΔΔCt method. Student's T-test was used to assess the expression difference between the two sample groups for statistical significance.

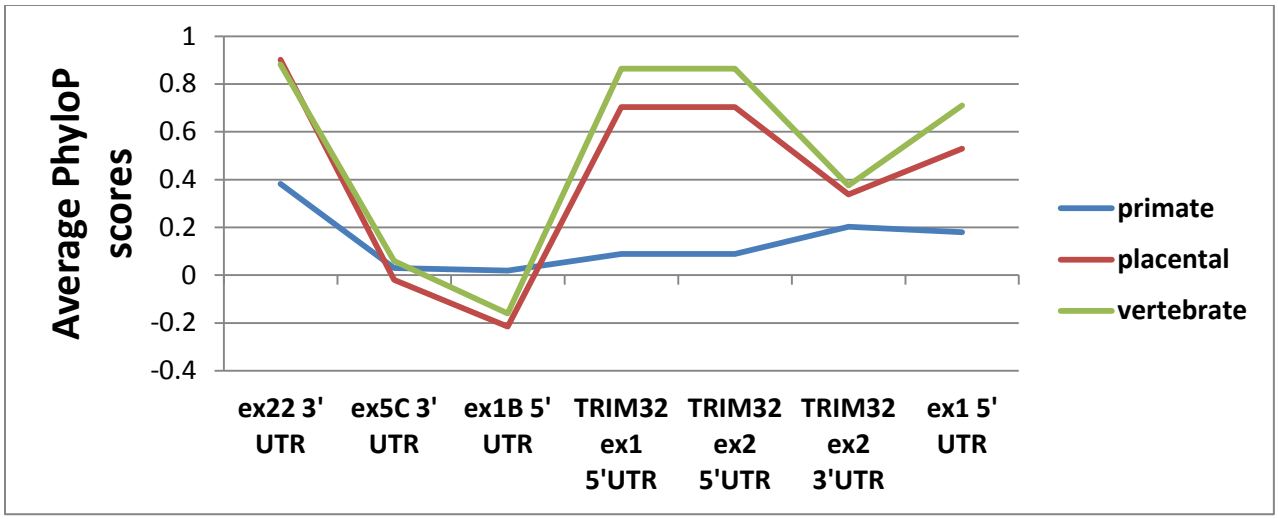

Figure S4. Conservation analysis of untranslated regions (UTRs) of *ASTN2* and *TRIM32* exons. The conservation was determined using average PhyloP scores calculated from nucleotide alignments between 46 vertebrate species including 23 placental mammals and eight primate species. Locations of the UTRs of different *ASTN2* transcript isoforms are labeled using the exon numbering in Figure 3A.

```
H.sapiens          MAAAGARLSPGPGSGLRGRPRLCFHPGPPPLLP--LLLLFLLLLP-PPPLLAGATAAAS- 56
M.mulatta          MAAAGARLSPGPGSGLRGRPRLRFHPGPPPLPP--LLLLFLLLLP-PPPLLAGATAAAS- 56
P.troglodytes      LPTASEPISPGHHSVLRVQPRICFHPGPPPLLP--LLLLLFLLAKQPPPLLAGATAAAS- 57
E.caballus         ------------------------------------------------------------
M.musculus         MAAAGARRSPGRGLGLRGRPRLGFHPGPPPPPPPPLLLLFLLLLP-PPPLLAGATAAAAS 59
C.familiaris       ------------------------------------------------------------
G.gallus           ------------------------------------------------------------
X.tropicalis       ------------------------------------------------------------

H.sapiens          REPDSPCRLKTVTVSTLPALRESDIGWSGARAGAGAGTGAG---------AAAAAASPGS 107
M.mulatta          REPDSPCRLKTVTVSTLPALRESDIGWSGARAGAGSGTGAG---------AAAAAS-PGS 106
P.troglodytes      REPDSPCRLKTVTVSTLPALRESDIGWSGARAGAGAGTGAG---------AAAAAS-PGS 107
E.caballus         ------------------------------------------------------------
M.musculus         REPDSPCRLKTVTVSTLPALRESDIGWSGARTGAAAGAGAGTGAGAGAAAAAASAASPGS 119
C.familiaris       ------------------------------------------------------------
G.gallus           ------------------------------------------------------------
X.tropicalis       ------------------------------------------------------------

H.sapiens          PGSAGTAAESRLLLFVRNELPGRIAVQDDLDNTELPFFTLEMSGTAADISLVHWRQQWLE 167
M.mulatta          PGSAGTAAESRLLLFVRNELPGRIAVQDDLDNTELPFFTLEMSGTAADISLVHWRQQWLE 166
P.troglodytes      PGSAGTAAESRLLLFVRNELPGRIAVQDDLDNTELPFFTLEMSGTAADISLVHWRQQWLE 167
E.caballus         ----------------------------------------EMSGTAADISLVHWRQQWLE 20
M.musculus         AGSAGTAAESRLLLFVRNELPGRIAVQDDLDNTELPFFTLEMSGTAADISLVHWRQQWLE 179
C.familiaris       ------------------------------------------------------------
G.gallus           -------------------------------------------MSGTVADISLVHWKQQWLE 19
X.tropicalis       ------------------------------------------------------------

H.sapiens          NGTLYFHVSMSSSGQLAQATAPTLQEPSEIVEEQMHILHISVMGGLIALLLLLLVFTVAL 227
M.mulatta          NGTLYFHVSMSSSGQLAQATAPTLQEPSEIVEEQMHILHISVMGGLIALLLLLLVFTVAL 226
P.troglodytes      NGTLYFHVSMSSSGQLAQATAPTLQEPSEIVEEQMHILHISVMGGLIALLLLLLVFTVAL 227
E.caballus         NGTLYFHVSMSSSGQLARATAPTLQEPSEIVEEQMHILHISVMGGLIALLLLLLVFTVAL 80
M.musculus         NGTLYFHVSMSSSGQLAQATAPTLQEPSEIVEEQMHILHISVMGGLIALLLLLLVFTVAL 239
C.familiaris       ---------MSSAGQLARATAPTLQEPSEIVEEQLHILHISVMGGLIALLLLLLVFTVAL 51
G.gallus           NGTLYFHVSMSSAEQLSRATPPSLQEPSEIVEEQMHILHISVMGGLIALLLLLLVFTVAL 79
X.tropicalis       ------------------------------------------------------------

H.sapiens          YAQRRWQKRRRIPQKSASTEATHEIHYIPSVLLGPQARESFRSSRLQTHNSVIGVPIRET 287
M.mulatta          YAQRRWQKRRRIPQKSASTEATHEIHYIPSVLLGPQARESFRSSRLQTHNSVIGVPIRET 286
P.troglodytes      YAQRRWQKRRRIPQKSASTEATHEIHYIPSVLLGPQARESFRSSRLQTHNSVIGVPIRET 287
E.caballus         YAQRRWQKRRRIPQKSASTEATHEIHYIPSVLLGPQARESFRSSRLQAHNSVIGVPIRET 140
M.musculus         YAQRRWQKRRRIPQKSASTEATHEIHYIPSVLLGPQARESFRSSRLQTHNSVIGVPIRET 299
C.familiaris       YAQRRWQKRRRVPQKSASTEATHEIHYIPSVLLGPQARESFRASRLQAHNSVIGVPIRET 111
G.gallus           YAQRRWHKRRRIPQKSASTEATHEIHYIPSVLLGPQARESFRNSRLQAHNSVIGVPIRET 139
X.tropicalis       ------------------------------------------------------------

H.sapiens          PILDDYDCEEDEEPPRRANHVSREDEFGSQVTHTLDSLGHPGEEKVDFEKK---------- 338
M.mulatta          PILDDYDCEEEEEPPRRANHVSREDEFGSQVTHTLDSLGRPGEEKVDFEKKAAAE--VTQ 344
P.troglodytes      PILDDYDCEEDEEPPRRANHVSREDEFGSQVTHTLDSLGHPGEEKVDFEKKAAAE--ATQ 345
E.caballus         PILDDYDYEEEEDPPRRANHVSREDEFGSQVTHTLDSLGRPGEEKGDFGKK---------- 191
M.musculus         PILDDYDYEEEEEPPRRANHVSREDEFGSQMTHALDSLGRPGEEKVEFEKKAAAE--ATQ 357
C.familiaris       PILDDYDYEDDEDLPRRTNHVSREDEFGSQVTHTLDSLGRPGEEKGDFEKKAAAE--ATQ 169
G.gallus           PILDDYEDEEEEETPQRAEPSTREDEFGSQVTRTLESLGRGGEEKPDYEKKGVAAPEVTQ 199
X.tropicalis       ------------------------------------------------------------

H.sapiens          ------------------------------------------------GGISFGRAKGTSGSE 353
M.mulatta          ETVESLMQKFKESFRANTPIEIGQLQPALRS-TSAGKRKRRSKSRGGISFGRAKGTSGSE 403
P.troglodytes      ETVESLMQKFKESFRANTPIEIGQLQPPLRS-TSAGKRKRRSKSRGGISFGRAKGTSGSE 404
E.caballus         ------------------------------------------------GGISFGRTKGMSGSE 206
M.musculus         ETVESLMQKFKESFRANTPVEIGQLQPASRSSTSAGKRKRRNKSRGGISFGRTKGTSGSE 417
C.familiaris       ETVESLMQKFKESFRANTPIEIGQLQPAPRR-ASAGRRKRRSKSRGGISFGRTKGTSGSE 228
G.gallus           ETVESLMQKFKESFRTNTPIEIGQLQPALRS-TSVGRRKRRSRPRGGIGFGRAKGNSGSE 258
X.tropicalis       ------------------------------------------------------------

H.sapiens          ADDETQLTFYTEQYRSRRRSKGLLKSPVNKTALTLIAVSSCILAMVCGSQMSCPLTVKVT 413
M.mulatta          ADDETQLTFYTEQYRSRRRSKGLLKSPVNKTALTLIAVSSCILAMVCGSQMSCPLTVKVT 463
P.troglodytes      ADDETQLTFYTEQYRSRRRSKGLLKSPVNKTALTLIAVSSCILAMVCGSQMSCPLTVKVT 464
E.caballus         ADDETQLTFYTEQYRSRRRSKGLLKSPVNKTALTLIAVSSCILAMVCGSQMSCPLTVKVT 266
M.musculus         ADDETQLTFYTEQYRSRRRSKGLLKSPVNKTALTLIAVSSCILAMVCGNQMSCPLTVKVT 477
C.familiaris       ADDETQLTFYTEQYRSRRRSKGLLKSPVNKTALTLIAVSSCILAMVCGTQMSCPLTVKVT 288
G.gallus           ADDETQLTFYTEQYRSRRRSKGSMKSPVNKTALTLIAVSSCILAMVCGSQLSCPLTVKVT 318
X.tropicalis       ------------------------------------------------VVCGSQLSCPLTVKVT 16
                                                                    :***.*:********
```

```
H.sapiens          LHVPEHFIADGSSFVVSEGSYLDISDWLNPAKLSLYYQINATSPWVRDLCGQRTTDACEQ 473
M.mulatta          LHVPEHFIADGSSFVVSEGSYLDISDWLNPAKLSLYYQINATSPWVRDLCGQRTTDACEQ 523
P.troglodytes      LHVPEHFIADGSSFVVSEGSYLDISDWLNPAKLSLYYQINATSPWVRDLCGQRTTDACEQ 524
E.caballus         LHVPEHFIADGSSFVVSEGSYLDISDWLNPAKLSLYYQINATSPWVRDLCGQRTTDACEQ 326
M.musculus         LHVPEHFIADGSSFVVSEGSYLDISDWLNPAKLSLYYQINATSPWVRDLCGQRTTDACEQ 537
C.familiaris       LHVPEHFIADGSSFVVSEGSYLDVSDWLNPAKLSLYYQINATSPWVRDLCGQRTTDACEQ 348
G.gallus           LHVPEHFIADGSSFVVSEGSYLDISDWLNPAKLSLYYQINATSPWVRDLCGQRTTDACEQ 378
X.tropicalis       LHVPEHFIADGSSFVISEGSYLDVSDWLNPAQLSLFYQVNSSSPWVQDLCGQRTTDACEQ 76
                   ***************:*******:*******:***:**:*::*****:**************

H.sapiens          LCDPETGECSC--HEGYAPDPVHRHLCVRSDWGQSEGPWPYTTLERGYDLVTGEQAPEKI 531
M.mulatta          LCDPETGECSC--YEGYAPDPVHRHLCVRSDWGQSEGPWPYTTLERGYDLVTGEQAPEKI 581
P.troglodytes      LCDPETGECSC--HEGYAPDPVHRHLCVRSDWGQSEGPWPYTTLERGYDLVTGEQAPEKI 582
E.caballus         LCDPETGECSC--HEGYAPDPVHRHLCVRSDWGQSEGPWPYTTLERGYDLVTGEQAPEKI 384
M.musculus         LCDPDTGECSC--HEGYAPDPVHRHLCVRSDWGQSEGPWPYTTLERGYDLVTGEQAPEKI 595
C.familiaris       LCDPETGECSC--HEGYAPDPVHRHLCVRSDWGQSEGPWPYTTLERGYDLVTGEQAPEKI 406
G.gallus           LCDQETGQGTCGWETDYAPDPMHRHLCVRTDWGHSEGPWPYTTLERGYDLVTGEQAPEKI 438
X.tropicalis       ICEPETGECSC--HEGYSPDATHRHLCVRSDWGQSQGPWPYNTLERGYDLVTGEQAPEKI 134
                   :*: :**: :*    .*:**. *******:***:*:*****.****************

H.sapiens          LRSTFSLGQGLWLPVSKSFVVPPVELSINPLASCKTDVLVTEDPADVR-EEAMLSTYFET 590
M.mulatta          LRSTFSLGQGLWLPVSKSFVVPPVELSINPLASCKTDVLVTEDPADVR-EEAMLSTYFET 640
P.troglodytes      L----SLGQGLWLPVSKSFVVPPVELSINPLASCKTDVLVTEDPADVR-EEAMLSTYFET 637
E.caballus         LRSTFSLGQGLWLPVSKSFVVPPVELSINPLASCKTDVLVTEDPADVR-EEAMLSTYFET 443
M.musculus         LRSTFSLGQGLWLPVSKSFVVPPVELSINPLASCKTDVLVTEDPADVR-EEAMLSTYFET 654
C.familiaris       LRSTFSLGQGLWLPVSKSFVVPPVELSINPLASCKTDVLVTEDPADVR-EEAMLSTYFET 465
G.gallus           LRSTYSLGQGLWLPVSKSFVVPPVELSINPLASCKTDVLVTEDPADVR-EEAMLSTYFET 497
X.tropicalis       LRSTYSLGQGIWLPVSKSFVVPPVELSINPLASCNTDVLVTEDPADVRRENALLATPFES 194
                   *    *****:**********************:************* *:*:*:* **:

H.sapiens          INDLLSSFGPVRDCSRNNGGCTRNFKCVSDRQVDSSGCVCPEELKPMKDGSGCYDHSKGI 650
M.mulatta          INDLLSSFGPVRDCSRNNGGCTRNFKCVSDRQVDSSGCVCPEELKPMKDGSGCYDHSKGI 700
P.troglodytes      INDLLSSFGPVRDCSRNNGGCTRNFKCVSDRQVDSSGCVCPEELKPMKDGSGCYDHSKGI 697
E.caballus         INDLLSSFGPVRDCSRNNGGCTRNFKCVSDRQVDSSGCVCPEELRPMKDGSGCYDHSKGI 503
M.musculus         INDLLSSFGPVRDCSRNNGGCTRNFKCVSDRQVDSSGCVCPEELKPMKDGSGCYDHSKGI 714
C.familiaris       INDLLSSFGPVRDCSRNNGGCTRNFKCVSDRQVDSSGCVCPEELRPMKDGSGCYDHSKGI 525
G.gallus           INDLLSSFGPVRDCSRNNGGCTRNFKCVSDRQVDSTGCVCPEDLRPMKDGTGCYDYSKGI 557
X.tropicalis       VDDLLSSFGPVRDCSLGNGGCSRNFKCVSERKIDSTGCVVP--LFLIRAGAGHPRWPGGP 252
                   ::************** .****:*******:*::**:*** *  *  :: *:*    . *


H.sapiens          DCSDGFNGGCEQLCLQQTLPLPYDATSSTIFMFCGCVEEYKLAPDGKSCLMLSDVCEGPK 710
M.mulatta          DCSDGFNGGCEQLCLQQTLPLPYDATSSTIFMFCGCVEEYKLAPDGKSCLMLSDVCEGPK 760
P.troglodytes      DCSDGFNGGCEQLCLQQTLPLPYDATSSTIFMFCGCVEEYKLAPDGKSCLMLSDVCEGPK 757
E.caballus         DCSDGFNGGCEQLCLQQTLPLPYDATSSTIFMFCGCVEEYKLAPDGKSCLMLSDVCEGPK 563
M.musculus         DCSDGFNGGCEQLCLQQTLPLPYDTTSSTIFMFCGCVEEYKLAPDGKSCLMLSDVCEGPK 774
C.familiaris       DCSDGFNGGCEQLCLQQTLPLPYDATSSTIFMFCGCVEEYKLAPDGKSCLMLSDVCEGPK 585
G.gallus           DCSDGFNGGCEQLCLQQTLPLPHDPSSSTIFMFCGCVEEYKLAPDGKSCLMLSDICEGPK 617
X.tropicalis       TGTEGRGNAMSVLGCSQVPLPCYCVTSLPVNMGCSCVEEYRLAPDGRSCLMLTDICEGSK 312
                     ::* ... . *  .*.    :  :* .: * *.*****:*****:*****:*:***.*

H.sapiens          CLKPDSKFNDTLFGEMLHGYNNRTQHVNQGQVFQMTFRENNFIKDFPQLADGLLVIPLPV 770
M.mulatta          CLKPDSKFNDTLFGEMLHGYNNRTQHVNQGQVFQMTFRENNFIKDFPQLADGLLVIPLPV 820
P.troglodytes      CLKPDSKFNDTLFGEMLHGYNNRTQHVNQGQVFQMTFRENNFIKDFPQLADGLLVIPLPV 817
E.caballus         CLKPDSKFNDTLFGEMLHGYNNRTQHVNQGQVFQMTFRENNFIKDFPQLADGLLVIPLPV 623
M.musculus         CLKPDSKFNDTLFGEMLHGYNNRTQHVNQGQVFQMTFRENNFIKDFPQLADGLLVIPLPV 834
C.familiaris       CLKPDSKFNDTLFGEMLHGYNNRTQHVNQGQVFQMTFRENNFIKDFPQLADGLLVIPLPV 645
G.gallus           CLKSDAKFNDTLFGEMLHGYNNRSQHVNQGQVFQMTFRENNFIKDFPQLADGLLVIPLPV 677
X.tropicalis       CLRPEDRLNDTLFGEMLHGYDNKTQKVNQGRVFQMSFRENNFIKDFPQLADGLLVIPLPV 372
                   **:.: ::*************:*::*:****:****:************************

H.sapiens          EEQCRGVLSEPLPDLQLLTGDIRYDEAMGYPMVQQWRVRSNLYRVKLSTITLAAGFTNVL 830
M.mulatta          EEQCRGVLSEPLPDLQLLTGDIRYDEAMGYPMVQQWRVRSNLYRVKLSTITLSAGFTNVL 880
P.troglodytes      EEQCRGVLSEPLPDLQLLTGDIRYDEAMGYPMVQQWRVRSNLYRVKLSTITLSAGFTNVL 877
E.caballus         EEQCRGVLSEPLPDLQLLTGDIRYDEAMGYPMVQQWRVRSNLYRVKLSTITLSAGFTNVL 683
M.musculus         EEQCRGVLSEPLPDLQLLTGDIRYDEAMGYPMVQQWRVRSNLYRVKLSTITLSAGFTNVL 894
C.familiaris       EEQCRGVLSEPLPDLQLLTGDVRYDEAMGYPMVQQWRVRSNLYRVKLSTITLSAGFTNVL 705
G.gallus           EEQCRGVLSEPLPDLQLLTGDIKYDEAMGYPMVQHWRVRSNLYHVKLSSITLSSGFANIL 737
X.tropicalis       EEQCRGVLSEPRPDLQQLTGDVRYNEAMGYPMVQQWRVRSNLYKVKLSSISLSTGFANVL 432
                   *********** **** ****::*:*********:********:****:*:*::**:*:*

H.sapiens          KILTKESSREELLSFIQHYGSHYIAEALYGSELTCIIHFPSKKVQQQLWLQYQKETTELG 890
M.mulatta          KILTKESSREELLSFIQHYGSHYIAEALYGSELTCIIHFPSKKVQQQLWLQYQKETTELG 940
P.troglodytes      KILTKESSREELLSFIQHYGSHYIAEALYGSELTCIIHFPSKKVQQQLWLQYQKETTELG 937
E.caballus         KILTKESSREELLSFIQHYGSHYIAEALYGSELTCIIHFPSKKVQQQLWLQYQKETTELG 743
M.musculus         KILTKESSRDELLSFIQHYGSHYIAEALYGSELTCIIHFPSKKVQQQLWLQYQKETTELG 954
C.familiaris       KILTRESSRDELLSFIQHYGSHYVAEALYGSELTCLIHFPSKKVQQQLWLQYQKETTELG 765
G.gallus           KILNKDSSREELLSFIQQFGSHYIAEALYGSEFSCTIHFPSKKVQQQLWLQYQKETTELG 797
X.tropicalis       KILSPQSSREDLLNVLHLYGSHYISEALYGSELTCTIRFPSKKVQQQLWLQYQKETTELG 492
                   ***. :***::**..:: :****::********:* *:**********************
```

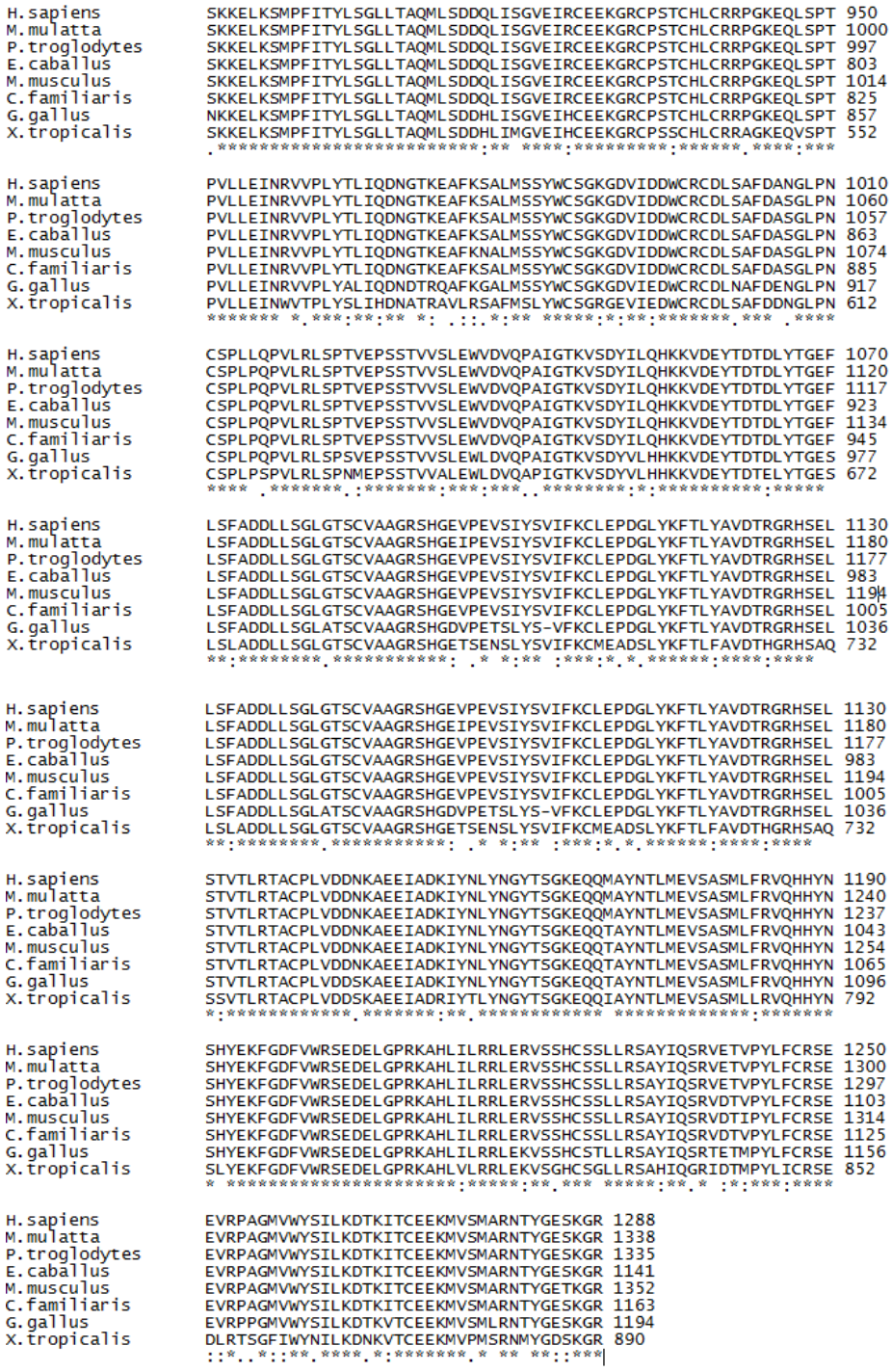

Figure S5. Amino acid alignment and protein conservation analysis of ASTN2. Eight protein sequences from 1:1 orthologs of human ASTN2 (ENSP00000354504): macaque (ENSMMUP00000039777), chimpanzee (ENSPTRP00000036394), horse (ENSECAP00000003457), mouse (ENSMUSP00000065786), dog (ENSCAFP00000005232), chicken (ENSGALP00000011381) and clawed frog (ENSXETP00000047008), were downloaded from Ensembl. "*" indicates identical residues in all sequences, ":" and "." denote conserved substitutions or semi-conserved substitutions respectively.

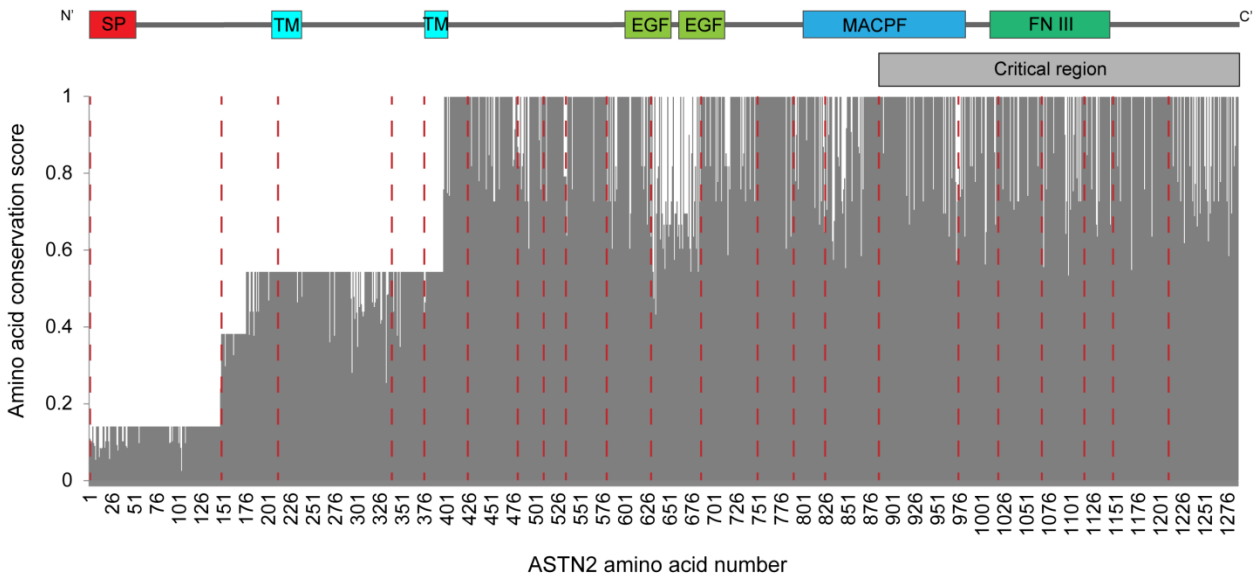

Figure S6. Quantification of residue conservation profile of ASTN2 (ENSP00000354504). Conservation score was calculated based on alignment in Figure S5 using the Scorecons server.(9) The known protein domains and features of ASTN2 are shown in the schematic illustration above the chart. The signal peptide (SP) is marked red, trans-membrane (TM) domains are marked light blue, EGF/laminin superfamily (EGF) domains are marked light green, the MACPF domain is marked blue and the fibronectin domain type 3 (FN III) is marked dark green. The critical region with enrichment of exonic deletions affecting multiple *ASTN2* isoforms in NDD cases from Figure 1 is shown in grey. Vertical red dashed lines correspond to the exon boundaries.

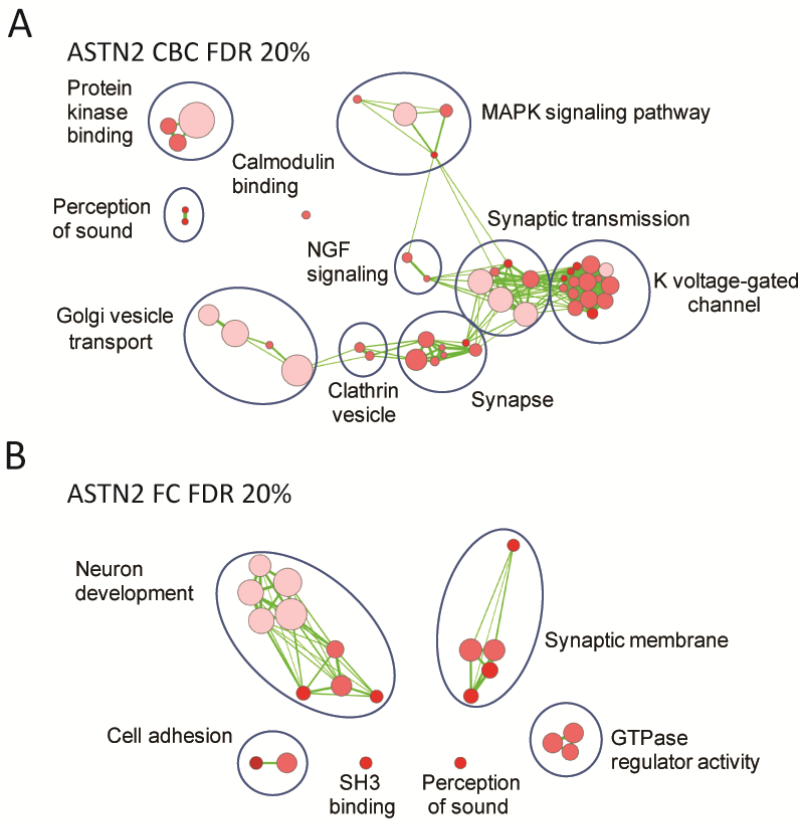

Figure S7. **Functional enrichment maps of genes co-expressed with *ASTN2*.** These maps depict the results of the gene set enrichment analysis of genes co-expressed with *ASTN2* in the frontal cortex (A) and the cerebellar cortex (B). Using the brain expression data from Brainspan,(10) the Pearson correlation coefficients were calculated between the expression levels of genes in the dataset and *ASTN2* in the CBC and FC regions. The top 500 most correlated genes for the two regions (Table S7) were used as inputs for the gene set enrichment analysis implementing manually curated gene ontology (GO) (org.Hs.eg.db, version 2.8.0) and pathways (downloaded from NCI, KEGG and Reactome websites) as previously described.(11, 12) Independently for the FC and the CBC, all expressed genes in each region were used as a background gene set and the statistical significance and false-discovery rate were calculated using the Fisher exact test and the Benjamini-Hochberg procedure respectively. The Cytoscape network software v.2.8.3 and the plugin Enrichment Map (version 1.2) were used to build the network. The functional enrichment maps were constructed using a false discovery rate cutoff of 20% and the gene-set clusters were manually identified and annotated (blue ovals) The node size is proportional to the gene-set size and the color corresponds to the odds ratio of enrichment (found in Table S7).

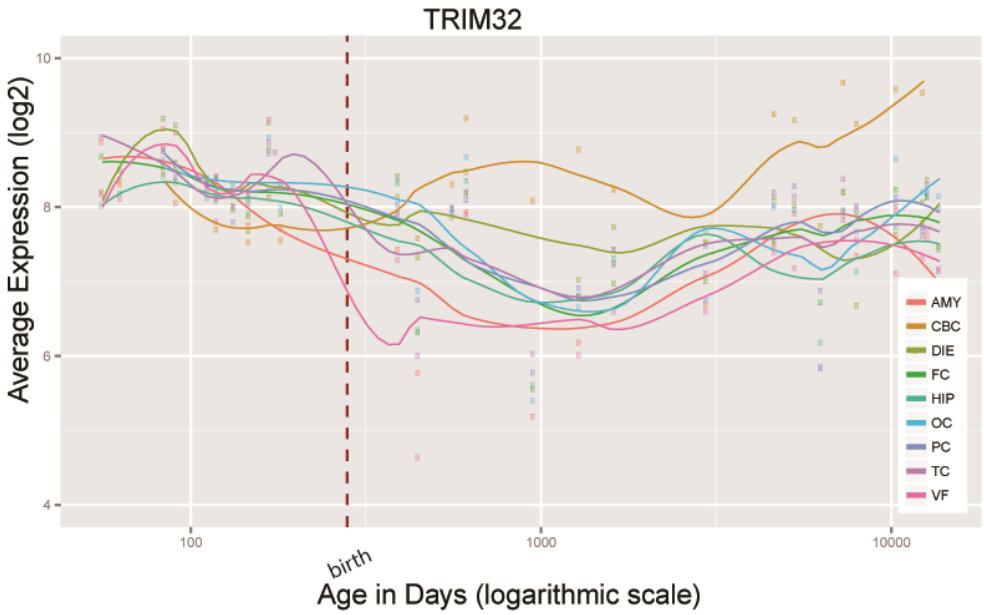

Figure S8. Expression profile of *TRIM32* across human brain development. The normalized expression data from the Brainspan database of *TRIM32* in amygdala (AMY), cerebellar cortex (CBC), diencephalon (DIE), frontal cortex (FC), hippocampus (HIP), occipital cortex (OC), parietal cortex (PC), temporal cortex (TC) and ventral forebrain (VF) was plotted across different developmental time-points.

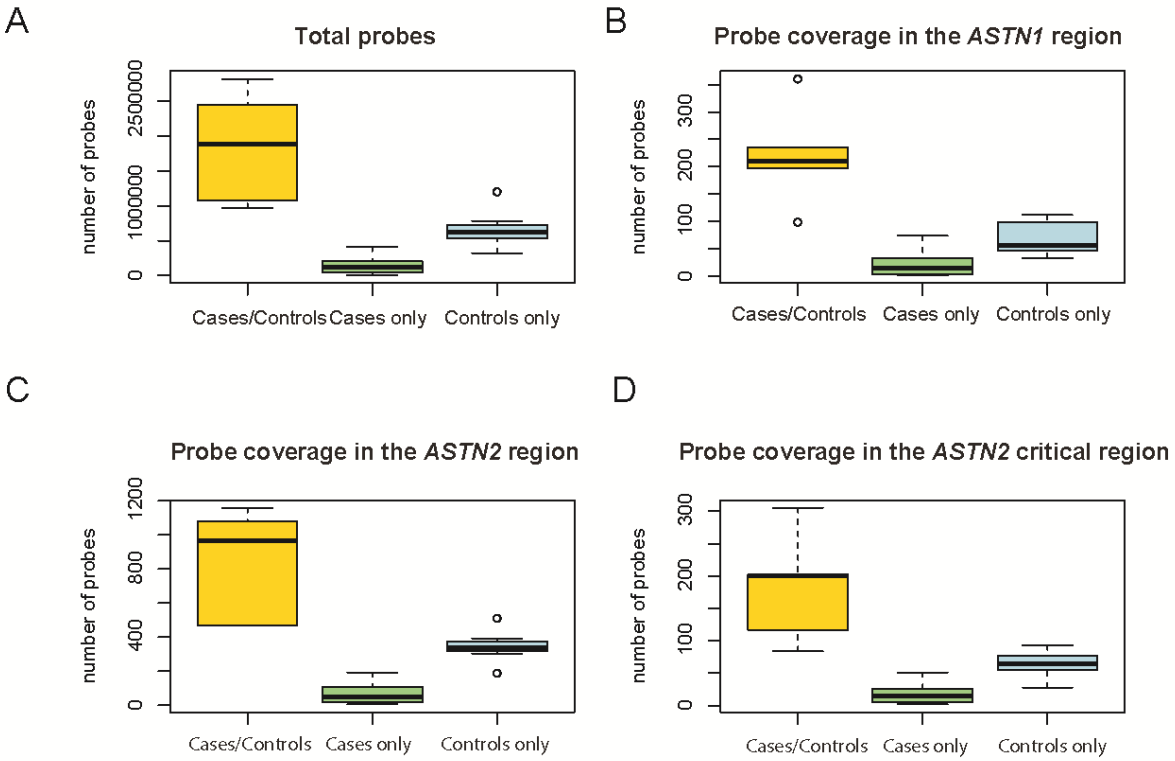

Figure S9. Comparisons of probe distributions of the microarray platforms used in this study. Box plots represent the distributions in microarray probe numbers across the genome (A) and within *ASTN1* (B), *ASTN2* (C) and the *ASTN2* critical region (D) (as highlighted by 3' end deletions affecting multiple isoforms) for the different platforms used in generating the CNV datasets utilized in this study (listed in Table S9). "Cases only" represent those microarray platforms utilized solely for CNV analysis of clinical samples (Cases in Table S9), "Controls only" represent microarray platforms from which CNV data was obtained for control individuals (Controls in Table S9) and "Cases/Controls" indicate microarray platforms from which CNV information was obtained for both case and control datasets (Cases, controls in Table S9). The plots reveal that the Control CNV dataset is generated from microarrays of resolutions significantly greater than or equal to those used to generate the Case CNV dataset.

Table S2. Control cohorts examined for CNVs at *ASTN2/TRIM32* and *ASTN1*

Control dataset	# Samples (males/females)[1]	Microarray platform	# exonic CNVs detected (at *ASTN2* / at *ASTN1*)[3]	Description of control cohort
Ontario ARCTIC	1,120 (629/491)	Affymetrix 500K	(0 / 0)	Zogopoulos *et al.* (13)
POPGEN	1,123 (623/500)	Affymetrix 6.0	(0 / 0)	Krawczak *et al.* (14)
Ottawa Heart Institute controls	1,234 (586/648)	Affymetrix 6.0	(0 / 0)	Stewart *et al.* (15)
HapMap (Phase 3)	1,056 (532/524)	Affymetrix 6.0	(1 loss / 0)	Altshuler *et al.* (16)
Starr County Diabetes study	1,794 (617/1,177)	Affymetrix 6.0	(0 / 0)	Below *et al.* (17)
Geneva NHS/HPFS Diabetes study	5,966 (2,608/3,358)	Affymetrix 6.0	(3 losses, 1 gain / 1 loss)	Qi *et al.* (18)
International Schizophrenia Consortium controls[2]	6,707 (N/A)	Affymetrix 6.0 & 5.0	(1 loss, 1 gain / 0)	ISC (19) Ernst *et al.* (20)
Ontario Population Genomics Platform (OPGP)	895 (489/406)	Affymetrix CytoHD	(1 loss / 0)	Costain *et al.* (21)
Population Diagnostics controls	1,000 (502/498)	Agilent 1M	(2 losses / 0)	Prasad *et al.* (22)
EDIC Diabetes study	1,422 (748/674)	Illumina 1M	(1 loss, 1 gain / 0)	Paterson *et al.* (23)
Wellcome Trust (WTCCC) controls	4,826 (2,412/2,414)	Illumina 1M	(4 losses / 0)	Rucker *et al.* (24)
SAGE consortium controls	1,287 (383/904)	Illumina 1M	(1 gain / 0)	Bierut *et al.* (25)
Health, Aging, and Body Composition (Health ABC) Study	2,566 (1,233/1,333)	Illumina 1M-Duo	(0 / 0)	Coviello *et al.* (26)
KORA	1,775 (855/920)	Illumina Omni 2.5M	(2 losses / 0)[4]	Verhoeven *et al.* (27)
COGEND	1,213 (498/715)	Illumina Omni 2.5M	(0 / 0)	Bierut *et al.* (28)
Shaikh *et al.*	2,026 (922/1,104)	Illumina 550K	(1 loss, 1 gain / 0)	Shaikh *et al.*(29)
Cooper *et al.* (HGDP)	984 (641/343)	Illumina 650Y	(0 / 0)	Cooper *et al.* (30)
Cooper *et al.* (London)	760 (384/376)	Illumina 550K	(0 / 0)	Cooper *et al.* (30)
Cooper *et al.* (FHCRC)	1,429 (0/1,429)	Illumina 610K Quad	(0 / 0)	Cooper *et al.* (30)
Cooper *et al.* (NINDS)	668 (N/A)	Illumina 550K & 317K	(0 / 0)	Cooper *et al.* (30)
Cooper et al. (PARC – CAP & PRINCE)	936 (N/A)	Illumina 317K	(1 loss / 0)	Cooper *et al.* (30)
Cooper et al. (PARC2 – CAP2 & PRINCE2)	766 (N/A)	Illumina 610K Quad	(0 / 0)	Cooper *et al.* (30)
Cooper *et al.* (inChianti)	695 (291/404)	Illumina 550K	(1 loss/ 0)	Cooper *et al.* (30)
Fernandez et al. controls	1,131 (N/A)	Illumina 1M-Duo	(0 / N/A)	Fernandez *et al.* (7)

Vrijenhoek et al. controls	706 (N/A)	Illumina 550K	(0 / N/A)	Vrijenhoek *et al*. (1)
Total	**44,085**		**(18 losses, 5 gains / 1 loss)**	

[1]This column displays total number of individuals and the split by gender for each control dataset. Gender information was not available (N/A) for some of the control datasets.

[2]The 2,090 and 1,171 Wellcome Trust control samples that were a part of the Cooper *et al*. control cohorts and Ernst *et al*. ISC control cohorts respectively were excluded to avoid double-counting of control individuals in this study.

[3]The coordinates of the specific CNVs detected in the control individuals are provided in Table S3.

[4] One individual in the KORA control cohort (CF14 in Figure 1 and Table S3) had an exonic deletion that overlapped *TRIM32* without affecting any exons of *ASTN2*.

Table S3: Coordinates of CNVs at *ASTN2/TRIM32* and *ASTN1* in control individuals

Sample	Control dataset	Gender	CNV coordinates (hg18)	CNV size	CNV type	Gene(s)
CM1	OPGP	male	chr9:118,361,183-118,524,134	162,952	loss	*ASTN2, TRIM32*
CM2	Cooper (Chianti)	male	chr9:118,497,798-118,530,393	32,596	loss	*ASTN2, TRIM32*
CM3	NHGRI	male	chr9:118,512,921-118,584,426	71,506	loss	*ASTN2*
CM4	NHGRI	male	chr9:118,523,510-118,589,363	65,854	loss	*ASTN2*
CM5	Population Diagnostics	male	chr9:118,532,361-118,539,807	7,447	loss	*ASTN2*
CM6	Population Diagnostics	male	chr9:118,555,826-118,657,525	101,700	loss	*ASTN2*
CM7	NHGRI	male	chr9:118,558,391-118,619,065	60,675	loss	*ASTN2*
CM8	WTCCC	male	chr9:118,651,651-118,879,149	227,499	loss	*ASTN2*
CM9	WTCCC	male	chr9:118,766,795-118,982,470	215,676	loss	*ASTN2*
CM10	Shaikh	male	chr9:118,828,951-118,927,398	98,448	loss	*ASTN2*
CF11	KORA	female	chr9:118,378,404-118,449,364	70,961	loss	*ASTN2*
CF12	WTCCC	female	chr9:118,471,284-118,510,780	39,497	loss	*ASTN2, TRIM32*
CF13	WTCCC	female	chr9:118,478,732-118,543,563	64,832	loss	*ASTN2 TRIM32*
CF14	KORA	female	chr9:118,501,096-118,507,817	6,722	loss	*TRIM32*
CF15	HapMap	female	chr9:118,514,521-118,531,233	16,713	loss	*ASTN2*
CF16	EDIC	female	chr9:118,543,563-118,621,959	78,397	loss	*ASTN2*
CF17	KORA	female	chr9:118,817,665-119,021,430	203,766	loss	*ASTN2*
CU18	ISC	N/A	chr9:118,327,376-118,510,195	182,820	loss	*ASTN2 TRIM32*
CU19	Cooper (ParcPrince)	N/A	chr9:118,765,546-118,839,405	73,860	loss	*ASTN2*
CM20	SAGE	male	chr9:118,619,065-118,859,862	240,798	gain	*ASTN2*
CM21	EDIC	male	chr9:118,651,651-119,045,245	393,595	gain	*ASTN2*
CM22	Shaikh	male	chr9:118,653,686-119,327,529	673,844	gain	*ASTN2*
CF23	NHGRI	female	chr9:118,479,893-118,508,356	28,464	gain	*ASTN2 TRIM32*
CU24	ISC	N/A	chr9:118,792,975-121,058,965	2,265,991	gain	*ASTN2*
CM25	NHGRI	male	chr1:174,960,861-175,139,720	178,860	loss	*ASTN1*

Table S4: *ASTN1* missense sequence variants in 338 Canadian ASD cases[1]

Position (hg19)	cDNA residue[2]	#AA[3]	#AB[3]	#BB[3]	Cases MAF[4]	dbSNP id	dbSNP MAF[4]	NHLBI MAF (%)[5]	Inh.[6]	Exon	Amino acid change	Protein domain	SIFT[7]	POLYPHEN[7]
chr1:176,999,962	1204(G>A)	0	1	337	0.0015	N/A	N/A	0	Paternal	4	R331Q	N/A	Damaging	Probably Damaging
chr1:176,913,063	2553(A>G)	0	1	337	0.0015	N/A	N/A	0	Paternal	14	I781V	N/A	Tolerated	Benign
chr1:176,863,877	2973(G>A)	0	1	337	0.0015	rs 201071031	0.0005	0	Paternal	17	G921S	Membrane attack complex/ perforin (MACPF) domain	Tolerated	Probably Damaging
chr1:176,852,074	3495(A>C)	0	1	337	0.0015	rs 151246825	0.0005	0.3953	Paternal	20	M1095L	Fibronectin type-III domain	Tolerated	Benign
chr1:176,833,571	3946(G>A)	0	1	337	0.0015	rs 148482637	N/A	0	Paternal	23	R1245H	N/A	Damaging	Probably Damaging
chr1:176,833,481	4036(C>T)	0	2	336	0.0030	rs 201817286	0.0005	0	Both Maternal	23	T1275M	N/A	Damaging	Probably Damaging
chr1:176,833,427	4090(A>G)	0	1	337	0.0015	rs 61756323	0.0005	0.4651	Maternal	23	E1293G	N/A	Damaging	Benign

[1] All exons of *ASTN1* (NM_004319.1) were sequenced via exome sequencing (in 306 unrelated ASD probands of European ancestry) or whole genome sequencing (in 32 unrelated ASD probands of European ancestry) and all rare (<1% in 1000 genomes) missense changes were recorded. Exome and WGS calls were confirmed using Sanger sequencing in the proband.
[2] The nucleotide change from the reference sequence and the cDNA position of this change are recorded in this column.
[3] These three columns denote the number of ASD cases with each genotype: homozygous for the reference allele (AA), heterozygous (AB) and homozygous for the minor allele (BB).
[4] The minor allele frequency (MAF) was calculated for our ASD case cohort and recorded as stated in the NCBI dbSNP database (build 137).
[5] The minor allele frequency (MAF) of these variants was determined from 4,300 individuals of European ancestry who were exome sequenced as part of the NIH National Heart, Lung and Blood Institute (NHLBI) Exome Sequencing Project (http://evs.gs.washington.edu/EVS/).
[6] Inheritance was determined via Sanger sequencing of both parents for all variants previously confirmed in probands.
[7] The effects of the resulting amino acid changes on protein function were predicted using SIFT Human Protein (http://sift.jcvi.org/) and Polyphen-2 (http://genetics.bwh.harvard.edu/pph2/).

Table S5: *ASTN2* missense sequence variants in 182 Canadian ASD cases[1]

Position (hg19)	cDNA residue[2]	#AA[3]	#AB[3]	#BB[3]	Cases MAF[4]	dbSNP id	dbSNP MAF[4]	NHLBI MAF (%)[5]	Inh.[6]	Exon	Amino acid change	Protein domain	SIFT[7]	POLYPHEN[7]
chr9:119,976,883	870(T>A)	0	1	181	0.0027	rs 139148246	0.0009	0.3953	Paternal	3	S257T	N/A	Damaging	Probably Damaging
chr9:119,858,397	1150(C>T)	0	1	181	0.0027	N/A	N/A	0	Paternal	4	S350L	N/A	Damaging	Probably Damaging
chr9:119,770,434	1476(G>T)	0	1	181	0.0027	rs 200108087	N/A	0	Maternal	6	V459L	N/A	Tolerated	Possibly Damaging
chr9:119,738,454	1638(A>T)	0	1	181	0.0027	rs 200986783	N/A	0	Maternal	8	T513S	N/A	Damaging	Benign
chr9:119,204,756	3522(C>A)	0	1	181	0.0027	rs 151278272	N/A	0.0698	Maternal	22	L1141M	N/A	Tolerated	Benign

[1] All exons of *ASTN2* (NM_014010.4) were sequenced via exome sequencing (in 159 unrelated ASD probands of European ancestry) or whole genome sequencing (in 23 unrelated ASD probands of European ancestry) and all rare (<1% in 1000 genomes) missense changes were recorded. Exome and WGS calls were confirmed using Sanger sequencing in the proband.
[2] The nucleotide change from the reference sequence and the cDNA position of this change are recorded in this column.
[3] These three columns denote the number of ASD cases with each genotype: homozygous for the reference allele (AA), heterozygous (AB) and homozygous for the minor allele (BB).
[4] The minor allele frequency (MAF) was calculated for our ASD case cohort and recorded as stated in the NCBI dbSNP database (build 137).
[5] The minor allele frequency (MAF) of these variants was determined from 4,300 individuals of European ancestry who were exome sequenced as part of the NIH National Heart, Lung and Blood Institute (NHLBI) Exome Sequencing Project (http://evs.gs.washington.edu/EVS/).
[6] Inheritance was determined via Sanger sequencing of both parents for all variants previously confirmed in probands.
[7] The effects of the resulting amino acid changes on protein function were predicted using SIFT Human Protein (http://sift.jcvi.org/) and Polyphen-2 (http://genetics.bwh.harvard.edu/pph2/).

Table S6: *TRIM32* missense sequence variants in 182 Canadian ASD cases[1]

Position (hg19)	cDNA residue[2]	#AA[3]	#AB[3]	#BB[3]	Cases MAF[4]	dbSNP id	dbSNP MAF[4]	NHLBI MAF[5]	Inh.[6]	Exon	Amino acid change	Protein domain	SIFT[7]	POLYPHEN[7]
chr9:119,460,622	762(C>T)	0	1	181	0.0027	rs147304059	0.0005	0	Paternal	2a	R201C	N/A	Damaging	Probably Damaging

[1] All exons of *TRIM32* (NM_012210.3) were sequenced via exome sequencing (in 159 unrelated ASD probands of European ancestry) or whole genome sequencing (in 23 unrelated ASD probands of European ancestry) and all rare (<1% in 1000 genomes) missense changes were recorded. Exome and WGS calls were confirmed using Sanger sequencing in the proband.
[2] The nucleotide change from the reference sequence and the cDNA position of this change are recorded in this column.
[3] These three columns denote the number of ASD cases with each genotype: homozygous for the reference allele (AA), heterozygous (AB) and homozygous for the minor allele (BB).
[4] The minor allele frequency (MAF) was calculated for our ASD case cohort and recorded as stated in the NCBI dbSNP database (build 137).
[5] The minor allele frequency (MAF) of these variants was determined from 4,300 individuals of European ancestry who were exome sequenced as part of the NIH National Heart, Lung and Blood Institute (NHLBI) Exome Sequencing Project (http://evs.gs.washington.edu/EVS/).
[6] Inheritance was determined via Sanger sequencing of both parents for all variants previously confirmed in probands.
[7] The effects of the resulting amino acid changes on protein function were predicted using SIFT Human Protein (http://sift.jcvi.org/) and Polyphen-2 (http://genetics.bwh.harvard.edu/pph2/).

Table S8. Cases from the DECIPHER database with exonic *ASTN2* CNVs

Decipher ID[1]	Sex	Location (hg18)	Size	CNV	Genes[2]	Phenotypes[3]	Inheritance
278838	F	chr9:116,513,630-120,227,136	3,713,507	loss	***ASTN2***, *TRIM32*, *PAPPA*, *TLR4*, *TNFSF15*, *TNFSF8*, *TNC*, *DEC1*	Intrauterine growth retardation, **Delayed speech and language development**, **Microcephaly**, Abnormality of the mouth, Abnormality of the heart	Paternal
257734	F	chr9:118,069,619-118,390,411	320,793	gain	***ASTN2***, *PAPPA*	Abnormality of the female genitalia	Inherited
259298	F	chr9:118,202,806-118,497,819	295,014	loss	***ASTN2***, *TRIM32*, *PAPPA*	Unknown	Inherited
268467	M	chr9:118,358,836-118,728,269	369,434	loss	***ASTN2***, *TRIM32*	**Intellectual disability**	Inherited
255906	M	chr9:118,420,432-118,497,819	77,388	loss	***ASTN2***, *TRIM32*	Unknown	Inherited
251779	M	chr9:118,440,935-118,584,415	143,481	loss	***ASTN2***, *TRIM32*	**Macrocephaly**, Strabismus, Epicanthus, Hyperextensible skin, **Intellectual disability**, Small nail	Unknown
277198	F	chr9:118,450,834-118,640,978	190,145	loss	***ASTN2***, *TRIM32*	Unknown	Unknown
264662	F	chr9:118,458,944-118,875,161	416,218	gain	***ASTN2***, *TRIM32*	Strabismus, **Intellectual disability**	Inherited
275492	M	chr9:118,459,293-118,558,331	99,039	loss	***ASTN2***, *TRIM32*	**Global developmental delay**, Postnatal **microcephaly**, Failure to thrive, High forehead, Fine hair, Prominent ears	Inherited
261647	M	chr9:118,474,920-118,643,716	168,797	loss	***ASTN2***, *TRIM32*	Abnormality of the face, **Intellectual disability**	Inherited
262543	M	chr9:118,479,880-118,514,893	35,014	loss	***ASTN2***, *TRIM32*	**Cognitive impairment**	Inherited
254230	M	chr9:118,572,595-118,667,250	94,656	loss	*ASTN2*	Unknown	Unknown
256747	F	chr9:118,608,198-118,669,889	61,692	loss	*ASTN2*	**Hyperactivity**, Dysarthria, **Intellectual disability**, Constipation	Inherited
271050	M	chr9:118,627,555-118,727,883	100,329	loss	*ASTN2*	**Delayed speech & language development, Intellectual disability**	Inherited
271612	F	chr9:118,728,270-118,934,998	206,729	gain	*ASTN2*	**Mild Intellectual disability**, Obesity	Unknown
258065	M	chr9:118,775,810-118,858,800	82,991	gain	*ASTN2*	Low-set ears, Leukodystrophy	*De novo*
255473	M	chr9:118,934,968-119,903,304	968,337	gain	*ASTN2*, *TLR4*	**Macrocephaly**, Hypertension, **Intellectual disability**, Obesity, Cerebellar vermis hypoplasia, Pericarditis, Sleep disturbance	Inherited

[1] The DECIPHER database ((https://decipher.sanger.ac.uk) was inspected (access date: November 15, 2013) for individuals with CNVs smaller than 6 Mb that were exonic to *ASTN2*. As indicated in this table, there are 17 such individuals and 11 of the 13 (85%) with information available about the reasons for referral for clinical microarray testing reported one or more neurodevelopmental disorder (NDD) traits. The majority of the *ASTN2* CNVs reported in DECIPHER (11/17) overlapped the 3' end of the gene and affected the shorter isoforms of the gene and/or *TRIM32*. We were able to obtain additional information about case 251779 (patient 18 in this study).

[2] Genes overlapped by the CNV. Bolded *ASTN2* indicates that multiple isoforms of the gene are overlapped by the CNV.

[3] Clinical information obtained from the DECIPHER database. Bolded terms represent neurodevelopmental disorder (NDD) traits.

[4] Information about inheritance of the CNV from the DECIPHER database. "Inherited" indicates inherited CNV (information about sex of transmitting parent was not available).

Table S9. Microarray probe coverage at *ASTN2/TRIM32* and *ASTN1*

Microarray platform	Datasets	Sites[1]	Total # array probes	# probes at *ASTN1*[2]	# probes at *ASTN2*[3]	# probes in critical region[4]
Affymetrix 250K	Cases	ITA	262,264	39	137	24
Agilent 44K	Cases	ITA	42,494	2	8	4
Agilent 4x44K ISCA	Cases	HSC,ITA,MC	42,869	3	18	5
Agilent 60K	Cases	ACH, BBG	62,976	2	6	2
Agilent 180K	Cases	ITA	170,334	29	73	21
Agilent 4x180K ISCA	Cases	CVH, ITA	174,200	16	62	18
Agilent 4x180K ISCA v2	Cases	HSC, MC	180,880	14	39	11
Agilent 244K	Cases	BCH	236,162	37	118	34
Agilent 400K	Cases	OUH	411,056	74	193	51
SignatureChipOS 105K	Cases	SG	95,933	5	18	5
SignatureChipOS v2 135K	Cases	SG	134,811	10	31	8
SignatureChipOS v3 135K	Cases	SG	138,466	31	99	28
Affymetrix SNP 6.0	Cases, controls	TCAG	1,880,794	235	969	202
Agilent 1M	Cases, controls	TCAG	967,029	197	470	117
Illumina 1M	Cases, controls	TCAG	1,072,820	99	468	84
Illumina Omni 2.5M	Cases, controls	TCAG	2,443,177	210	1,079	201
Affymetrix CytoScan HD	Cases, controls	TCAG	2,819,494	360	1,159	305
Affymetrix 500K	Controls	-	501,138	90	305	66
Affymetrix SNP 5.0	Controls	-	781,522	112	361	88
Illumina 317K	Controls	-	318,237	34	188	28
Illumina 550K	Controls	-	561,303	47	337	56
Illumina 610K Quad	Controls	-	620,901	47	330	55
Illumina 650Y	Controls	-	660,755	56	393	64
Illumina 1M-Duo	Controls	-	1,199,187	107	511	93

[1] Molecular diagnostic testing site of origin of cases. ACH, Alberta Children's Hospital; BBG, Brain and Body Genetic Resource Exchange (BBGRE); BCH, Boston Children's Hospital; CVH, Credit Valley Hospital; HSC, The Hospital for Sick Children; ITA, Italian diagnostic labs; MC, Mayo Clinic; OUH, Odense University Hospital; SG, Signature Genomics; TCAG, The Centre for Applied Genomics.
[2] Number of array probes within *ASTN1* (hg18 coordinates of chr1:175,096,826-175,400,647).
[3] Number of array probes within *ASTN2* (hg18 coordinates of chr9:118,227,325-119,217,138).
[4] Number of array probes within the critical region encompassing *TRIM32* and multiple isoforms of *ASTN2* (hg18 coordinates of chr9:118,227,325-118,503,400).

Table S10. RT-PCR and qRT-PCR primer sets used for *ASTN2* expression analysis

Primer name	Forward primer	Reverse primer	Accession number for transcript detected
ASTN2 **All isoforms pair1***	AGGGAGGCACTCAGAGCTAA	CCATCTGCTGCTCCTTTCCA	NM_014010, NM_198186, NM_198188, NM_198187, NM_001184734, NM_001184735
ASTN2 **All isoforms pair2***	ATCGATGACTGGTGCAGGTG	TACTGGAGGGCTCCACTGTT	NM_014010, NM_198186, NM_198188, NM_198187, NM_001184735
ASTN2 **Long isoform pair3***	CAGGTGGATTCCTCGGGATG	GCCGCCATTAAAGCCATCAG	NM_014010
ASTN2 **Long isoform pair4***	CTTTGTGCGTAACGAGCTGC	CTGTGCCAGACATCTCCAGG	NM_014010
ASTN2 **Shorter isoforms**	ATGAAGAAACCGACGCCCAT	ATCTGACAGCATCTGGGCTG	NM_198187, NM_198188, NM_001184734, NM_001184735
ASTN2 **Shorter isoforms pair5***	GCTCAAGTTCACTCAAGACCACA	ATCTGACAGCATCTGGGCTG	NM_198187, NM_198188, NM_001184734, NM_001184735
ASTN2 **NM_198186**	AGGCAGACTTGGGGAGTACA	ATCTGACAGCATCTGGGCTG	NM_198186
ASTN2 **NM_001184734 pair6**	ATCGATGACTGGTGCAGGTG	CAAGGCTGGTTCAATATCCGC	NM_001184734

*qRT-PCR primers

Supplementary References

1 Vrijenhoek, T., Buizer-Voskamp, J.E., van der Stelt, I., Strengman, E., Sabatti, C., Geurts van Kessel, A., Brunner, H.G., Ophoff, R.A. and Veltman, J.A. (2008) Recurrent CNVs disrupt three candidate genes in schizophrenia patients. *Am. J. Hum. Genet.*, **83**, 504-510.

2 Lionel, A.C., Crosbie, J., Barbosa, N., Goodale, T., Thiruvahindrapuram, B., Rickaby, J., Gazzellone, M., Carson, A.R., Howe, J.L., Wang, Z. *et al.* (2011) Rare copy number variation discovery and cross-disorder comparisons identify risk genes for ADHD. *Sci Transl Med*, **3**, 95ra75.

3 Bernardini, L., Alesi, V., Loddo, S., Novelli, A., Bottillo, I., Battaglia, A., Digilio, M.C., Zampino, G., Ertel, A., Fortina, P. *et al.* (2010) High-resolution SNP arrays in mental retardation diagnostics: how much do we gain? *Eur. J. Hum. Genet.*, **18**, 178-185.

4 Glessner, J.T., Wang, K., Cai, G., Korvatska, O., Kim, C.E., Wood, S., Zhang, H., Estes, A., Brune, C.W., Bradfield, J.P. *et al.* (2009) Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature*, **459**, 569-573.

5 Grozeva, D., Kirov, G., Ivanov, D., Jones, I.R., Jones, L., Green, E.K., St Clair, D.M., Young, A.H., Ferrier, N., Farmer, A.E. *et al.* (2010) Rare copy number variants: a point of rarity in genetic risk for bipolar disorder and schizophrenia. *Arch Gen Psychiatry*, **67**, 318-327.

6 Vulto-van Silfhout, A.T., Hehir-Kwa, J.Y., van Bon, B.W., Schuurs-Hoeijmakers, J.H., Meader, S., Hellebrekers, C.J., Thoonen, I.J., de Brouwer, A.P., Brunner, H.G., Webber, C. *et al.* (2013) Clinical significance of de novo and inherited copy-number variation. *Hum. Mutat.*, **34**, 1679-1687.

7 Fernandez, T.V., Sanders, S.J., Yurkiewicz, I.R., Ercan-Sencicek, A.G., Kim, Y.S., Fishman, D.O., Raubeson, M.J., Song, Y., Yasuno, K., Ho, W.S. *et al.* (2012) Rare copy number variants in tourette syndrome disrupt genes in histaminergic pathways and overlap with autism. *Biol Psychiatry*, **71**, 392-402.

8 Seno, M.M., Hu, P., Gwadry, F.G., Pinto, D., Marshall, C.R., Casallo, G. and Scherer, S.W. (2011) Gene and miRNA expression profiles in autism spectrum disorders. *Brain Res.*, **1380**, 85-97.

9 Valdar, W.S. (2002) Scoring residue conservation. *Proteins*, **48**, 227-241.

10 Kang, H.J., Kawasawa, Y.I., Cheng, F., Zhu, Y., Xu, X., Li, M., Sousa, A.M., Pletikos, M., Meyer, K.A., Sedmak, G. *et al.* (2011) Spatio-temporal transcriptome of the human brain. *Nature*, **478**, 483-489.

11 Pinto, D., Pagnamenta, A.T., Klei, L., Anney, R., Merico, D., Regan, R., Conroy, J., Magalhaes, T.R., Correia, C., Abrahams, B.S. *et al.* (2010) Functional impact of global rare copy number variation in autism spectrum disorders. *Nature*, **466**, 368-372.

12 Merico, D., Isserlin, R., Stueker, O., Emili, A. and Bader, G.D. (2010) Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One*, **5**, e13984.

13 Zogopoulos, G., Ha, K.C., Naqib, F., Moore, S., Kim, H., Montpetit, A., Robidoux, F., Laflamme, P., Cotterchio, M., Greenwood, C. *et al.* (2007) Germ-line DNA copy number variation frequencies in a large North American population. *Hum. Genet.*, **122**, 345-353.

14 Krawczak, M., Nikolaus, S., von Eberstein, H., Croucher, P.J., El Mokhtari, N.E. and Schreiber, S. (2006) PopGen: population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. *Community Genet*, **9**, 55-61.

15 Stewart, A.F., Dandona, S., Chen, L., Assogba, O., Belanger, M., Ewart, G., LaRose, R., Doelle, H., Williams, K., Wells, G.A. *et al.* (2009) Kinesin family member 6 variant Trp719Arg does not associate with angiographically defined coronary artery disease in the Ottawa Heart Genomics Study. *J. Am. Coll. Cardiol.*, **53**, 1471-1472.

16 Altshuler, D.M., Gibbs, R.A., Peltonen, L., Dermitzakis, E., Schaffner, S.F., Yu, F., Bonnen, P.E., de Bakker, P.I., Deloukas, P., Gabriel, S.B. *et al.* (2010) Integrating common and rare genetic variation in diverse human populations. *Nature*, **467**, 52-58.
17 Below, J.E., Gamazon, E.R., Morrison, J.V., Konkashbaev, A., Pluzhnikov, A., McKeigue, P.M., Parra, E.J., Elbein, S.C., Hallman, D.M., Nicolae, D.L. *et al.* (2011) Genome-wide association and meta-analysis in populations from Starr County, Texas, and Mexico City identify type 2 diabetes susceptibility loci and enrichment for expression quantitative trait loci in top signals. *Diabetologia*, **54**, 2047-2055.
18 Qi, L., Cornelis, M.C., Kraft, P., Stanya, K.J., Linda Kao, W.H., Pankow, J.S., Dupuis, J., Florez, J.C., Fox, C.S., Pare, G. *et al.* (2010) Genetic variants at 2q24 are associated with susceptibility to type 2 diabetes. *Hum. Mol. Genet.*, **19**, 2706-2715.
19 InternationalSchizophreniaConsortium (2008) Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature*, **455**, 237-241.
20 Ernst, C., Marshall, C.R., Shen, Y., Metcalfe, K., Rosenfeld, J., Hodge, J.C., Torres, A., Blumenthal, I., Chiang, C., Pillalamarri, V. *et al.* (2012) Highly penetrant alterations of a critical region including BDNF in human psychopathology and obesity. *Arch Gen Psychiatry*, **69**, 1238-1246.
21 Costain, G., Lionel, A.C., Merico, D., Forsythe, P., Russell, K., Lowther, C., Yuen, T., Husted, J., Stavropoulos, D.J., Speevak, M. *et al.* (2013) Pathogenic rare copy number variants in community-based schizophrenia suggest a potential role for clinical microarrays. *Hum. Mol. Genet.*, **22**, 4485-4501.
22 Prasad, A., Merico, D., Thiruvahindrapuram, B., Wei, J., Lionel, A.C., Sato, D., Rickaby, J., Lu, C., Szatmari, P., Roberts, W. *et al.* (2012) A discovery resource of rare copy number variations in individuals with autism spectrum disorder. *G3 (Bethesda)*, **2**, 1665-1685.
23 Paterson, A.D., Waggott, D., Boright, A.P., Hosseini, S.M., Shen, E., Sylvestre, M.P., Wong, I., Bharaj, B., Cleary, P.A., Lachin, J.M. *et al.* (2010) A genome-wide association study identifies a novel major locus for glycemic control in type 1 diabetes, as measured by both A1C and glucose. *Diabetes*, **59**, 539-549.
24 Rucker, J.J., Breen, G., Pinto, D., Pedroso, I., Lewis, C.M., Cohen-Woods, S., Uher, R., Schosser, A., Rivera, M., Aitchison, K.J. *et al.* (2013) Genome-wide association analysis of copy number variation in recurrent depressive disorder. *Mol. Psychiatry*, **18**, 183-189.
25 Bierut, L.J., Agrawal, A., Bucholz, K.K., Doheny, K.F., Laurie, C., Pugh, E., Fisher, S., Fox, L., Howells, W., Bertelsen, S. *et al.* (2010) A genome-wide association study of alcohol dependence. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 5082-5087.
26 Coviello, A.D., Haring, R., Wellons, M., Vaidya, D., Lehtimaki, T., Keildson, S., Lunetta, K.L., He, C., Fornage, M., Lagou, V. *et al.* (2012) A genome-wide association meta-analysis of circulating sex hormone-binding globulin reveals multiple Loci implicated in sex steroid hormone regulation. *PLoS Genet.*, **8**, e1002805.
27 Verhoeven, V.J., Hysi, P.G., Wojciechowski, R., Fan, Q., Guggenheim, J.A., Hohn, R., MacGregor, S., Hewitt, A.W., Nag, A., Cheng, C.Y. *et al.* (2013) Genome-wide meta-analyses of multiancestry cohorts identify multiple new susceptibility loci for refractive error and myopia. *Nat. Genet.*, **45**, 314-318.
28 Bierut, L.J., Madden, P.A., Breslau, N., Johnson, E.O., Hatsukami, D., Pomerleau, O.F., Swan, G.E., Rutter, J., Bertelsen, S., Fox, L. *et al.* (2007) Novel genes identified in a high-density genome wide association study for nicotine dependence. *Hum. Mol. Genet.*, **16**, 24-35.
29 Shaikh, T.H., Gai, X., Perin, J.C., Glessner, J.T., Xie, H., Murphy, K., O'Hara, R., Casalunovo, T., Conlin, L.K., D'Arcy, M. *et al.* (2009) High-resolution mapping and analysis of copy number variations in the human genome: a data resource for clinical and research applications. *Genome Res.*, **19**, 1682-1690.
30 Cooper, G.M., Coe, B.P., Girirajan, S., Rosenfeld, J.A., Vu, T.H., Baker, C., Williams, C., Stalker, H., Hamid, R., Hannig, V. *et al.* (2011) A copy number variation morbidity map of developmental delay. *Nat. Genet.*, **43**, 838-846.