Fig. S1A

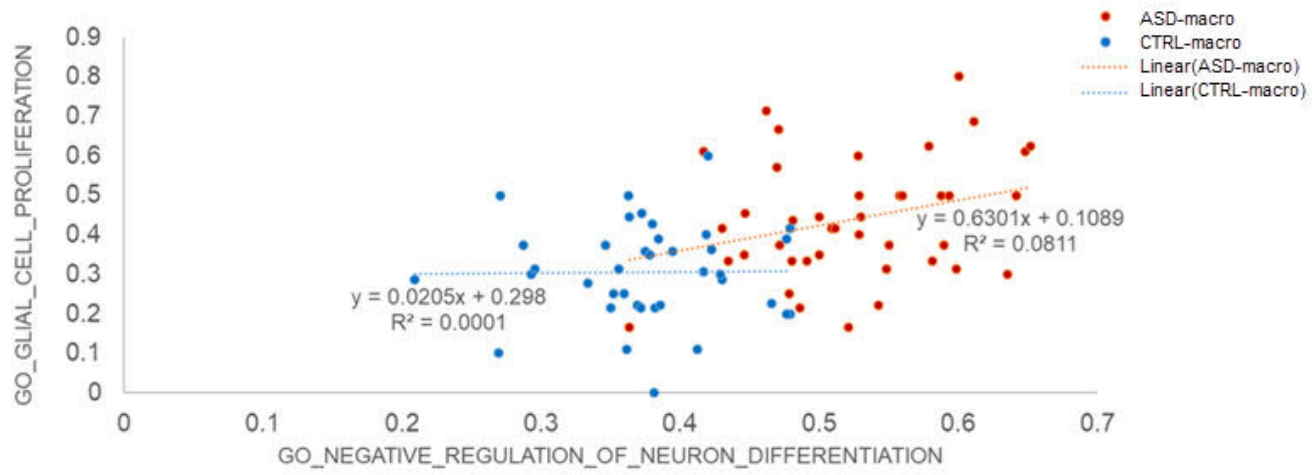


Fig. S1B

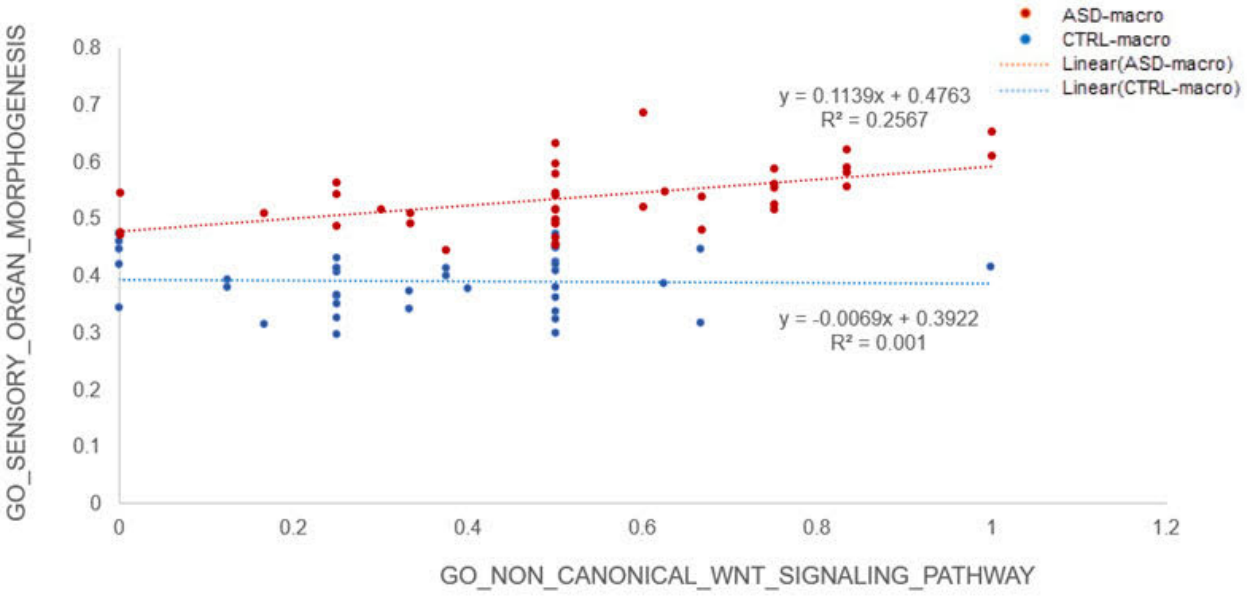


Fig. S1C

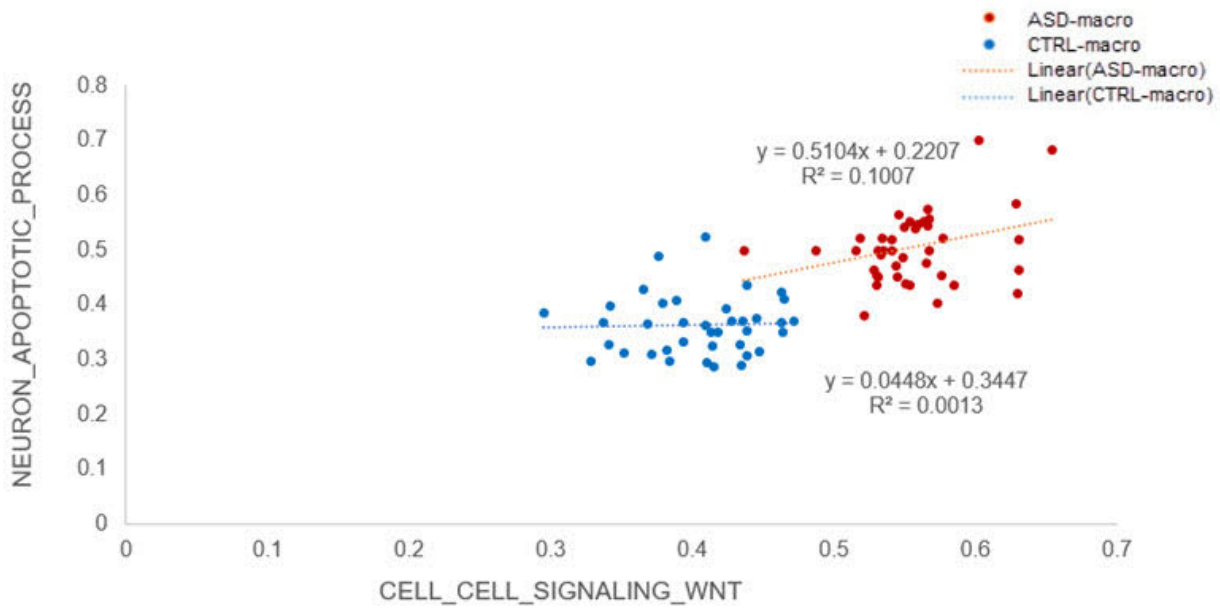


Figure S1. A-C. Additional examples of GO-GO correlation between sGOs and cGOs in ASD-macro samples. variation_rate$_{GO}$ of sGOs were ploted on X-axis, variation_rate$_{GO}$ of cGOs were ploted on Y-axis. Each dot represent one individual. Red color for ASD and blue for CTRL. Linear regression line was plotted for ASD and CTRL separately. Linear regression equation and variance explained by linear regeression was shawn on top of each regression line. A. Negative regulation of neuron differentiation vs. glial cell proliferation. B. Non canonical WNT signaling pathway vs. sensory organ morphogenesis. **(C)** Cell cell signaling WNT vs. neuron apoptotic process.

Fig.S2A.

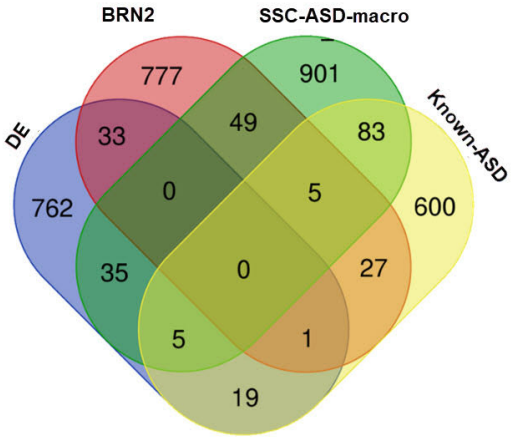


Fig.S2B.

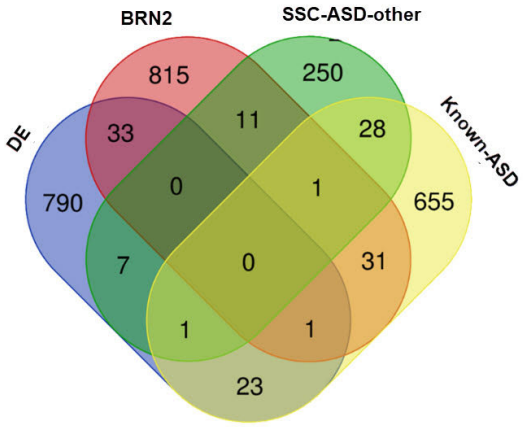


Fig.S2A. Venn-diagram for variations selected from SFARI-SSC ASD-macro samples (green circle), *BRN2* binding genes (red circle), differentially expressed (DE) genes (blue circle) and known ASD genes (yellow circle). **S2B**: variations selected from SFARI-SSC ASD-other samples and *BRN2* binding genes, differentially expressed (DE) genes and known ASD genes.

Fig.S3A

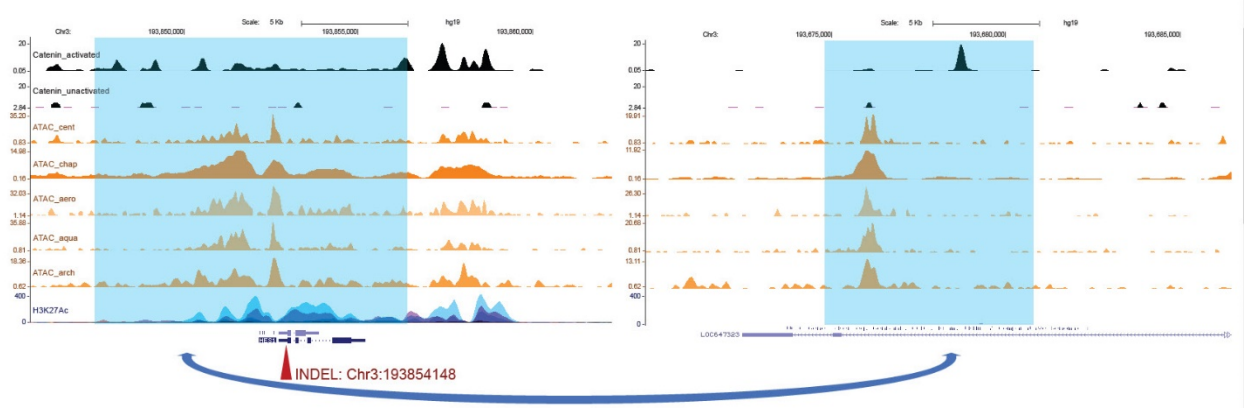
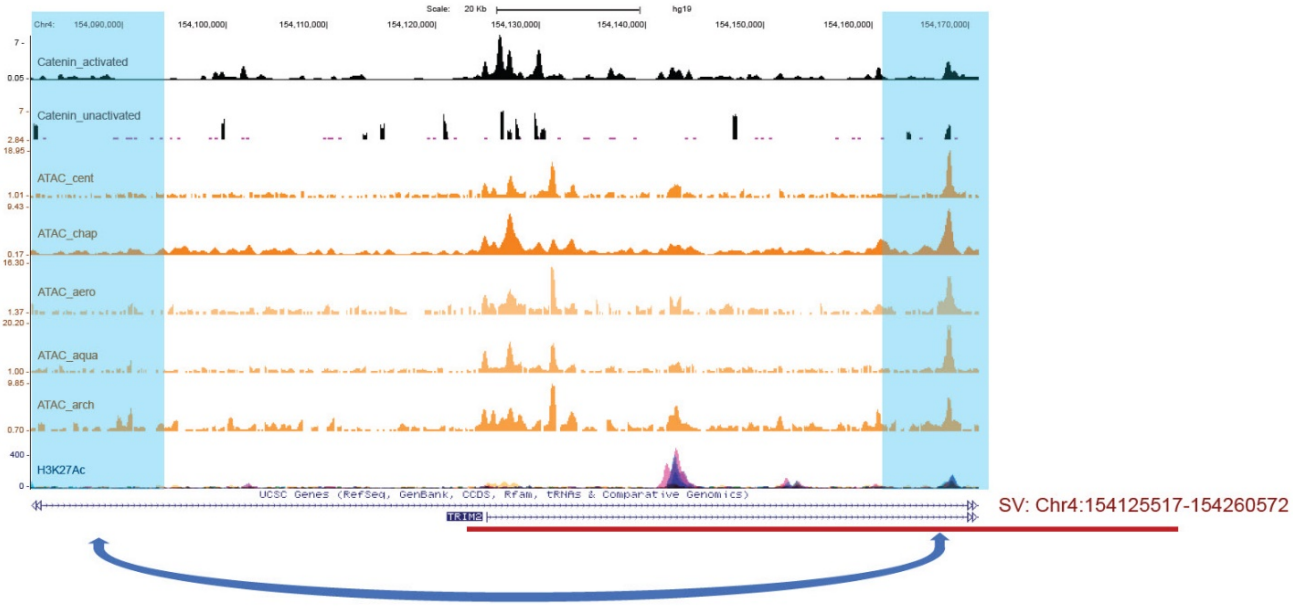
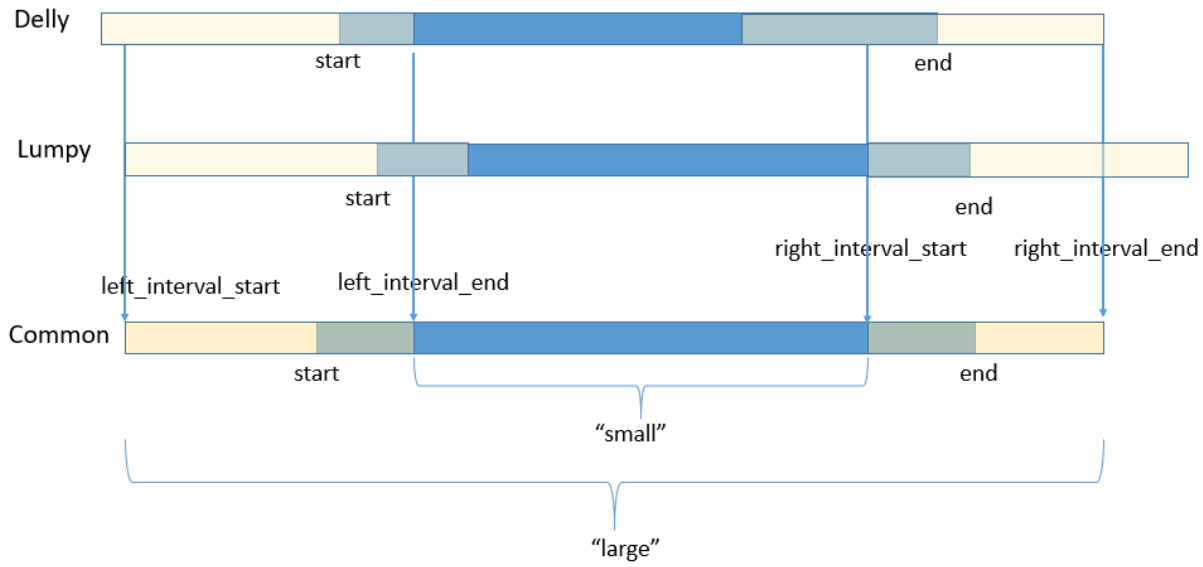


Fig.S3B



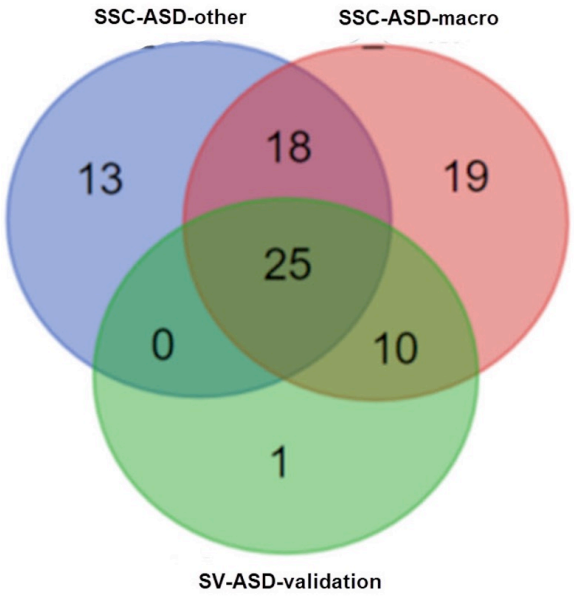
Supplementary figure 3A. INDEL on *HES1* (Chr3:193854148, red arrow pointed) was selected from SFARI ASD macrocephaly samples. It is on first exon of *HES1*, and close to *CTNNB1* peak (*WNT3A* activated, top black track) confirmed by ATACseq peaks (orange tracks). These peaks (including the gene and the INDEL) was in one DNA interval (blue shade, left), looping with another DNA interval on chromosome 3 (blue shade, right), which covered *CTNNB1* peak on gene *LOC647323*. **3B.** Example plot for SV (Chr4:154125517-154260572) on *TRIM2*, which overlapped with *CTNNB1* peak and HiC interval. This interval interacts with a distant interval overlapped with the 5' part of the gene *TRIM2* through DNA looping at NPC stage.

Fig.S4



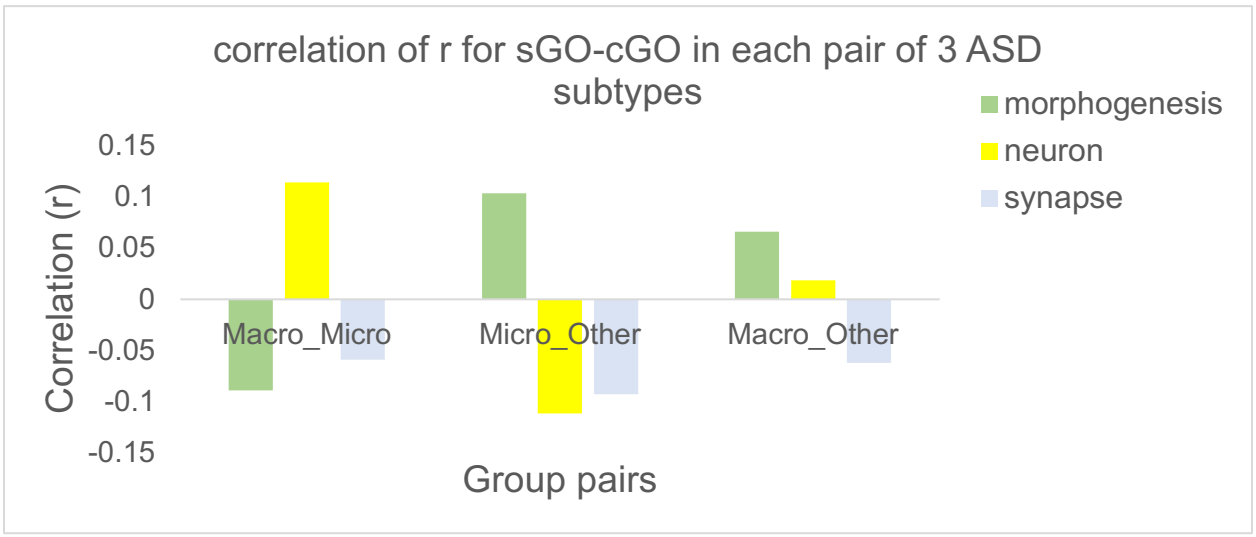
Supplementary figure 4. Collapse of deletions called by both *Lumpy* and *Delly* in validation dataset. Blue bars represent the deletion (from start to end breakpoint), yellow bars represent 95% confidence interval for start or end breakpoint. Common intervals (bottom track) were decided by overlap between intervals estimated by *Delly* (top track) and *Lumpy* (middle track). "Small" estimation of the deletion event used the right side of the common start interval and left side of the common end interval; "Large" estimation of the deletion event used the left side of the common start interval and right side of the common end interval.

Fig. S5



Supplementary figure 5. Venn diagram for TFs with regulation targets enriched in genes affected by ASD-macro specific SVs from validation dataset (green circle) and TFs with regulation targets enriched in genes carrying SFARI-SSC ASD-macro variations (red circle) and TFs with regulation targets enriched in genes carrying ASD-other variations (blue circle) based on *ChEA* analysis.

Fig. S6



Supplementary figure 6. Bar plot of spearman correlation for ranks of coefficient for sGO-cGO pairs between each pair of ASD subgroup. No significant result was detected.

Supplementary Methods

CTRL individual from NIST for SV analysis

To avoid bias introduced by the small control sample number (n=2) for matepair sequencing, we downloaded 3 CEPH samples with matepair sequencing data from National Institute of Standards and Technology (NIST) [62] and call deletions following a similar pipeline as processing samples in the replication dataset. After merging deletions called from these 3 samples with those called from our 2 control samples, we found in total 1321 deletions (extended control list), including 248 shared with our ASD deletions and 1073 control specific deletions (Table S8E). With the larger number of deletions from the new control dataset, the number of ASD specific deletions (n=406), intersected genes (n=229) and GO terms (Table S8E, 8F) enriched by these genes were largely conserved as using the 2 controls we sequenced, suggesting the bias caused by small number of controls was trivial. In this way, our control deletion list (n_{sample}=2) was used for subsequent analysis in this paper.