

Supplementary text

Supplementary Table S1 : selected positive and negative ASD genes

Our training consisted of known ASD-associated genes (from the high confidence genes in the Simons Foundation Autism Research Initiative (SFARI) Gene database (category 1, category 2, syndromic genes and the 65 genes established in the Sander study (sanders et al. 2015)) as positives and non-mental-health-related genes as negatives (selected from the list of non-ASD genes created by Krishnan et al (Krishnan et al. 2016).

Supplementary Figure S1. Heatmap visualisation of human brain spatiotemporal gene expression. Each cell show the expression level of a gene (row) in a specific brain region and development stage (column). The ASD risk and non-risk genes are indicated by red and blue rows, respectively. The number denote the ID of genes. The ASD and non ASD genes are denoted by red and blue rows, respectively. Different colors indicate brain regions. The ASD risk genes are expressed in a higher level in comparison to non-risk genes across developmental stages and brain regions. The intensity of the color represents the log2-transformed expression level.

CBC, cerebellar cortex; CB, cerebellum; VFC, ventrolateral prefrontal cortex; MC, motor cortex ; SC, motor-sensory cortex ; MCSC, primary motor-sensory cortex ; IPC, posteroventral (inferior) parietal cortex; PCx, parietal neocortex; HIP, hippocampus ; DTH, dorsal thalamus; TCx, temporal neocortex; MD, mediodorsal nucleus of thalamus; AC, auditory cortex (core); AMY, amygdaloid complex; STR, striatum; URL, upper (rostral) rhombic lip; OFC, orbital frontal cortex; Ocx, occipital neocortex; MGE, medial ganglionic eminence; CGE, caudal ganglionic eminence; LGE, lateral ganglionic eminence; STC, posterior (caudal) superior temporal cortex (area 22c); MFC, anterior (rostral) cingulate (medial prefrontal) cortex; VC, primary visual cortex (striate cortex, area V1/17); ITC, inferolateral temporal cortex (area TEv, area 20); DFC, dorsolateral prefrontal cortex.

Supplementary Figure S2: Confusion matrix for each model of machine learning

A confusion matrix consists of 4 different combinations having actual and predicted values. It is used to describe the performance of a classification model on a set of test data for which the true values are known. This helps to give information about whether the data is being correctly classified or not. ANN model have the best performance for classification (predicted the high number of correct true positive genes)

Supplementary table S2 : Gene risk score predicted from ANN model

Supplementary table S3 : GO enrichment analysis for predicted ASD risk genes

Supplementary table S4 : Functionnal clustering of GO terms