Microarray datasets (GSE26415) Two groups: ASD group (n=21) & control group (n=21) Step 1: data pre-processing Background correction, Quantile normalization, Filtering Step 2: dividing datasets Into a training dataset (n=26) & test dataset (n=16) Step 3: feature selection Identification of differentially expressed probes from the training dataset Moderated t-test with Bonferrini correction (ρ <0.05) Step 4b: supervised machine learning Step 4a: unsupervised (classification) machine learning (clustering) Machine learning algorithms: SVM, KNN, LDA Hirarchical clustering analysis on all samples (n=42) Training on the training dataset (n=26), and validation on the test dataset (n=16)

Supplementary Fig. 2. Diagram of the study design. ASD, autism spectrum disorder; SVM, support vector machine; KNN, K-nearest neighbor; LDA, linear discriminant analysis.