Supplementary materials

Supple. Figure. The prediction powers using all features with different sample sizes. In each sample size, we randomly sampled 500 times and modeling according to the procedure discussed in method 2.3 in each sample. It can be seen that the prediction power increases with the rising of sample size. However, when the sample size reaches 2000, the prediction power increases only moderately. We chose 5000 as our sample size. The mean prediction power of 5000 sample size is 0.774±0.006 .