

Difference and correlation

Introduction

Difference of mOTUs between HCs and SZs can obtain by Wilcoxon rank sum test and zero-inflated model. Wilcoxon rank sum test is much more generally, and zero-inflated model was used for data set with excess zeros. After that, p value was adjusted by q value method.

Correlation most often refers to how close two variables are to having a linear relationship with each other. Generally speaking, there are three commonly used correlation coefficients, Pearson's correlation coefficient, Spearman's rank correlation coefficient and Kendall tau rank correlation coefficient. Because Pearson's correlation coefficient is sensitive to the data distribution, we mostly use the other two methods to calculate the correlation of two variables.

Data info

171 samples collected from community and Research center for mental disorders, includes 81 healthy controls (HC) and 90 schizophrenes (SZ). In SZ, 49 patients are first-episode, and 41 patients are relapse. 360 mOTUs profile which represents bacteria abundance was calculated by analyzing the sequencing data. As the metagenomics data, mOTUs profile contains numerous zero. (file name ../../SZData/motu.txt)

The Positive and Negative Syndrome Scale (PANSS) is a medical scale used for measuring symptom severity of patients with schizophrenia. The patient is rated from 1 to 7 on 30 different symptoms based on the interview as well as reports of primary care hospital workers. It has 3 catalogs, positive scale (p1-p7 in our research), negative scale (m1-m7 in our research) and general psychopathology scale (g1-g16 in our research). The minimum score of PANSS total is 30, and the maximum score of PANSS is 210. It should be noted that 1 rather 0 is the lowest score for each item.

According to definition of PANSS, just SZs (90) did this scale. Except the original 30 items, our cooperating parties combined some items together as new items based on their features (file name ../../SZData/panss.txt).

Procedure

Difference

The purpose of calculating difference in SZs' mOTUs and HCs' mOTUs is to find mOTUs with enrichment. As for total mOTUs, the original method was using Wilcoxon rank sum test. If mOTUs mean of SZs is larger than HCs that means mOTUs enrich in SZs. Because of excess zeros in mOTUs

profile, zero-inflated model is fitted at present. In this model, state (0 or 1) represent predictors, each mOTU represent outcome. The abundance of mOTUs in each sample is 0, which means it is not present. If the count of mOTUs not present is smaller than 160 and larger than or equal with 34 (20% presence rate in 171 samples), I used zero-inflated model. If the count is smaller than 34, I used Wilcoxon rank sum test. Finally, the total p value gotten from these two methods were adjusted by q value. The statistics is larger than zero, means mOTU enrich in HCs, or it is smaller than zero, means mOTU enrich in SZs. The code of difference is named zero_state.R.

Correlation

As for part of mOTUs profile, matching with PANSS, and then filtering mOTUs profile with 20% occurrence rate cutoff, correlation of mOTUs and PANSS was calculated using Kendall tau rank correlation coefficient. In the heat map, “+” represents p value smaller than 0.05, “*” represents p value smaller than 0.01. Except the labels in heat map, mOTUs with enrichment and PANSS were clustered. The code of correlation and heat map were named by cor_motu_phe.R and heatmap.R respectively.

According to the heat map, there are two question. Firstly, the mOTUs with same enrichment were not clustered together in some SZs. The second, as I described before, PANSS scale is used for evaluating the degree of SZs, which should be positive correlated with SZs and negative correlated with HCs, but it is opposite in some SZs.