**Name: Le Chang**

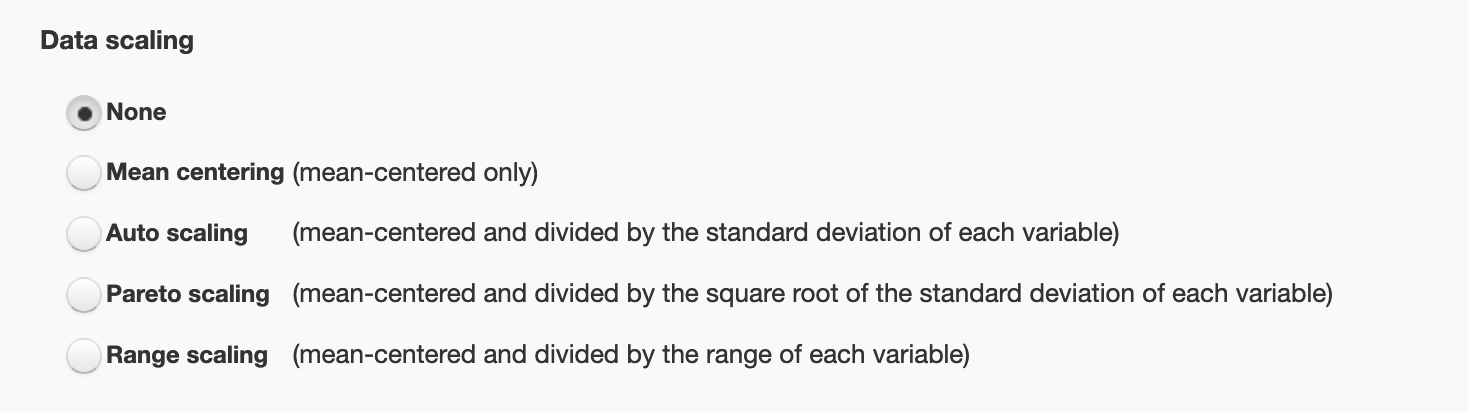
**ID: 260905699**

**Metaboanalyst**

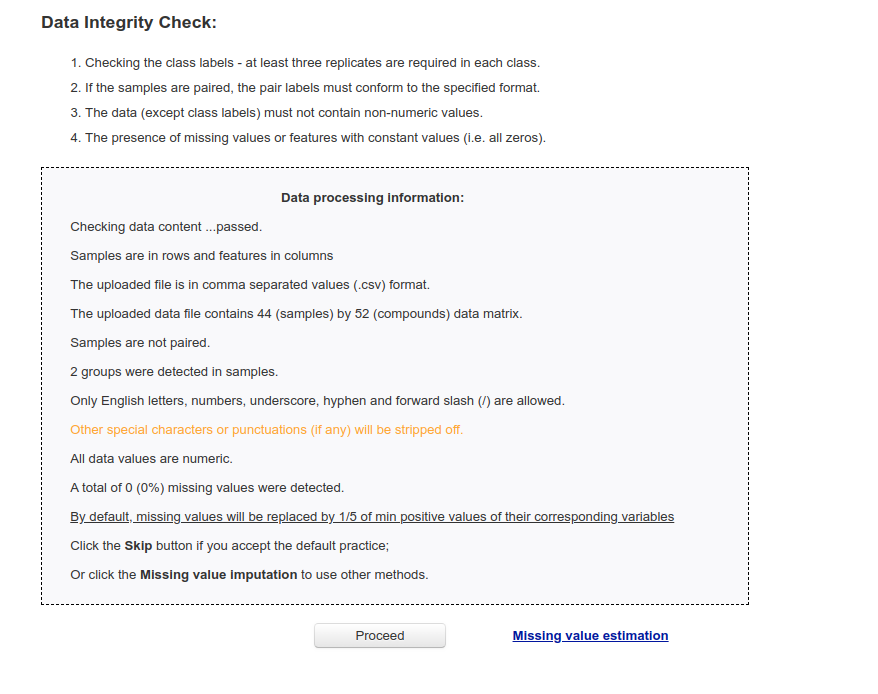
*GBA* mutations represent the greatest genetic risk factor for Lewy Body Diseases, both Parkinson’s Disease (PD) and Dementia with Lewy Bodies (DLB) (Clark et al., 2010; Tsuang et al., 2012). Heterozygous *GBA* mutations are found in ~25% of all DLB and >10% of PD patients (Clark et al., 2010; Tsuang et al., 2012). *GBA* encodes lysosomal acid-beta-glucosidase (GCase). GCase hydrolyzes glucosylceramides (GlcCers) to glucose and ceramides and glycosylsphingosine (GlcSph) to sphingosine. GCase activity is reduced to 58% of normal levels in *GBA* carriers with PD and 67% in idiopathic PD patients (wild-type *GBA*).

However, not all patients who carry PD-associated *GBA* mutations will develop PD. You have been given a metabolomics dataset performed on plasma of persons who carry a PD-associated *GBA* mutation and are either **assymptomatic** (carriers) or **symptomatic** (show the symptoms of PD). Using metaboanalyst, perform the following analyses and provide a screen shot of each analysis:

1. The metabolite abundances in these datasets are very different. First, scale your data. You have the following choices. Which did you choose and why?

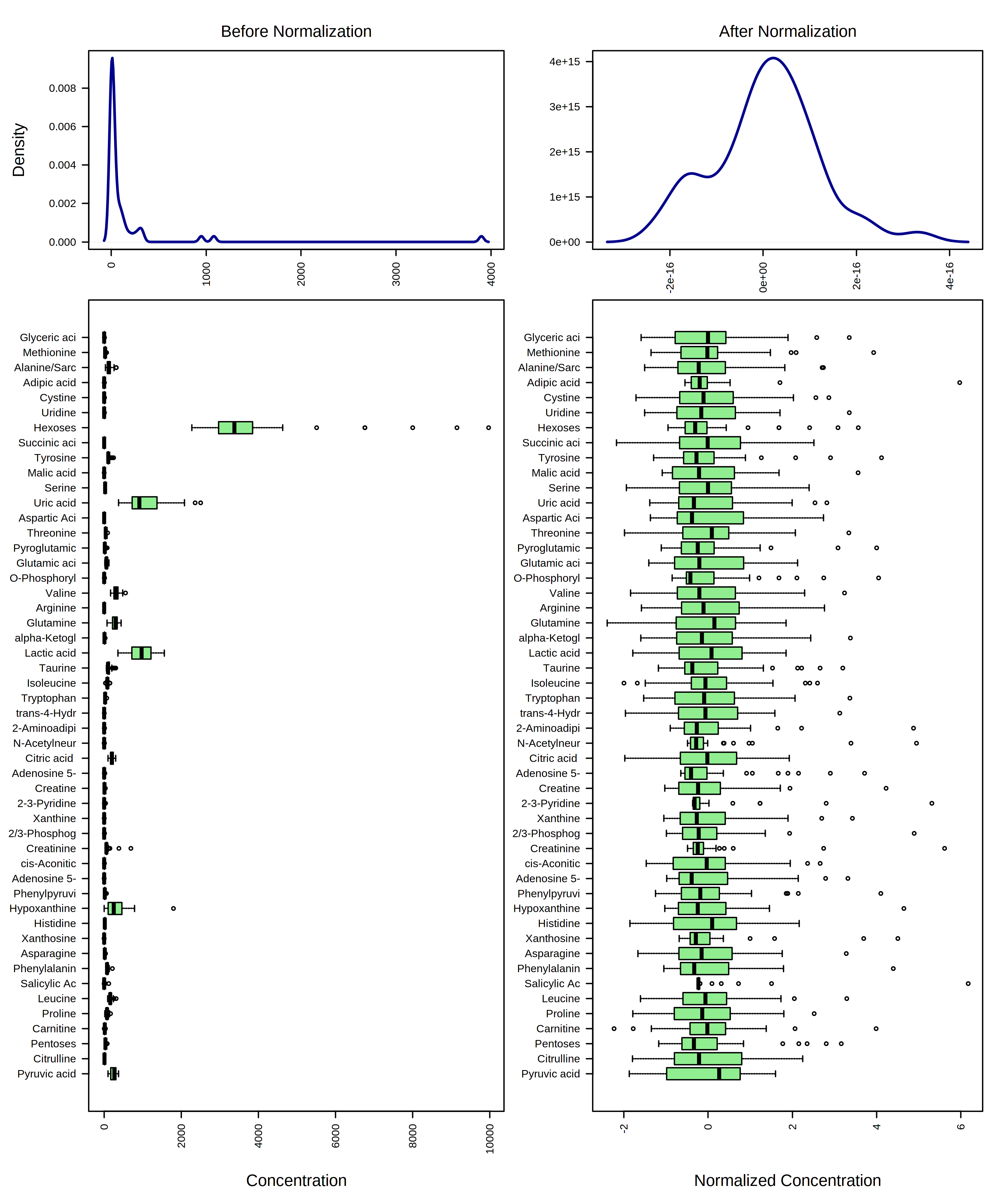


Step 1- Data integrity check:

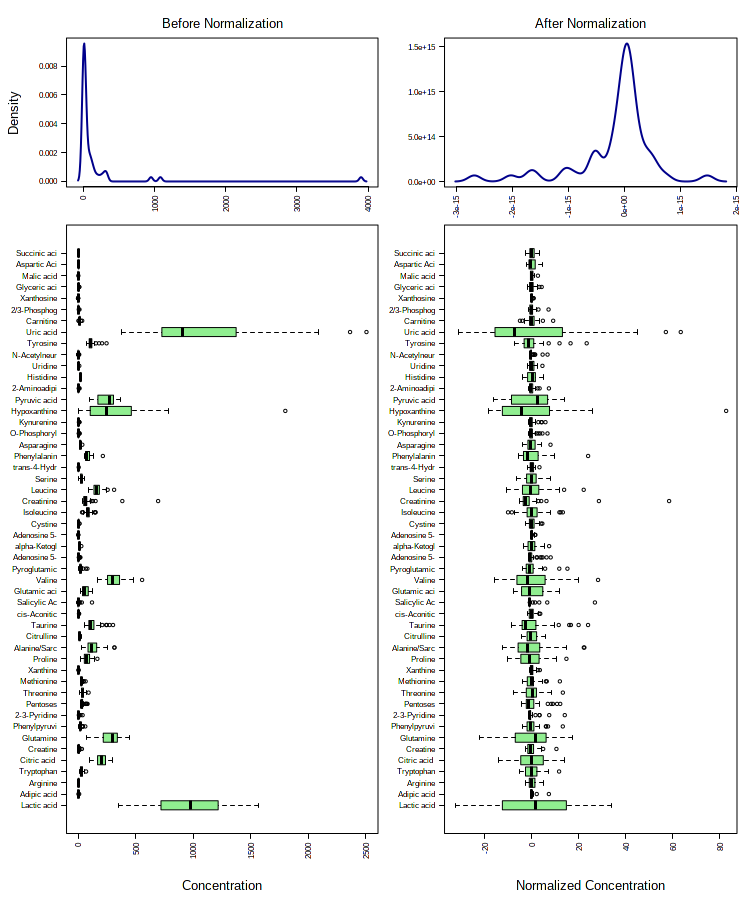


Step 2 - Scaling

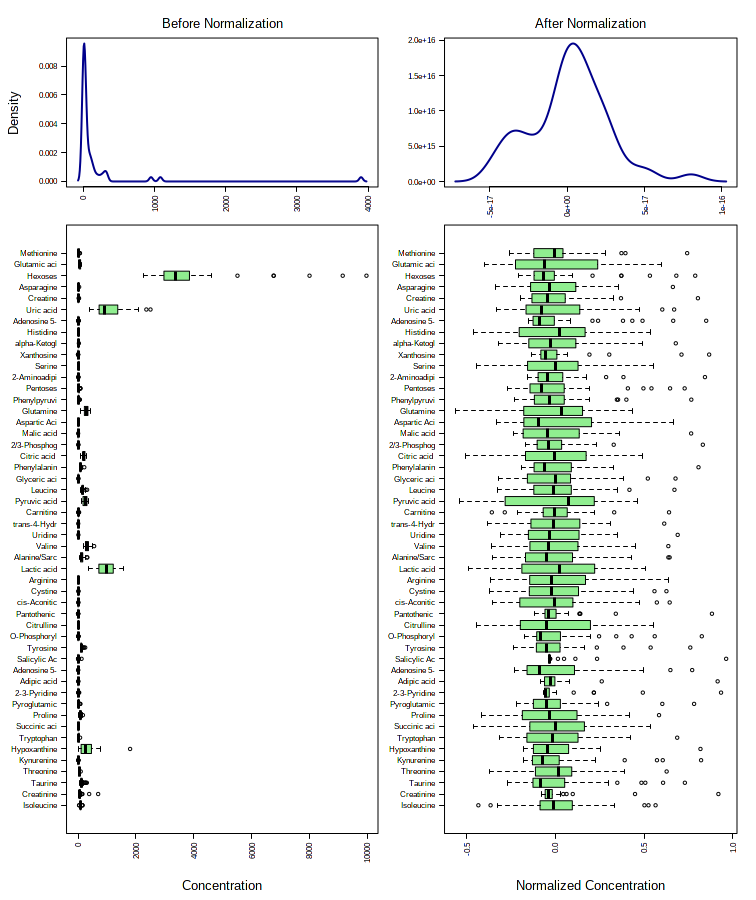
Method 1: Auto scaling:



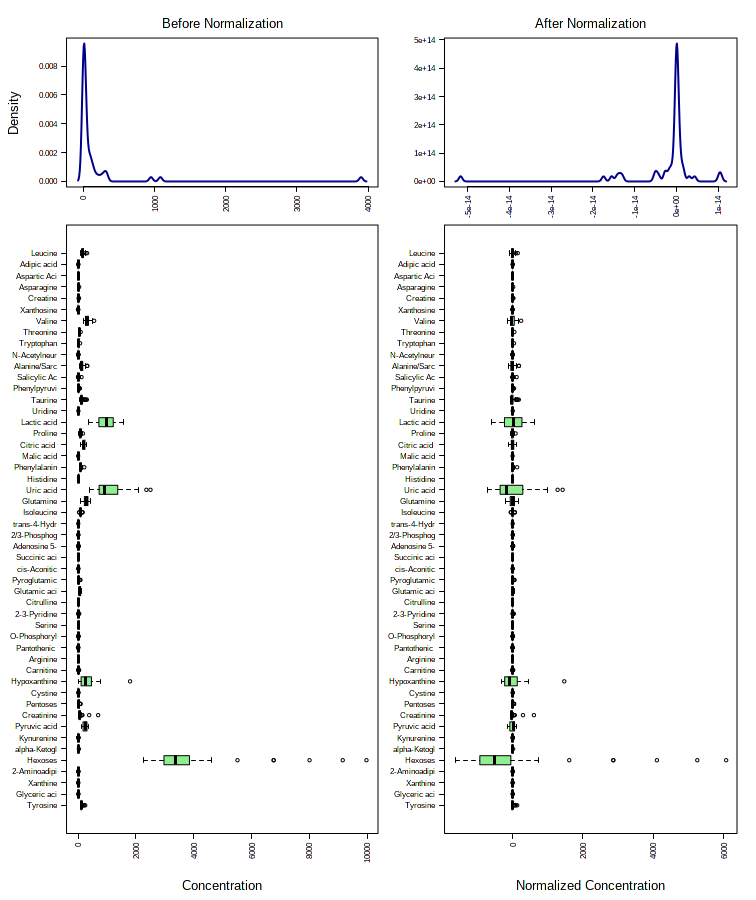
Method 2 - Pareto scaling



Method 3 - Range scaling



Method 4 - Mean centering

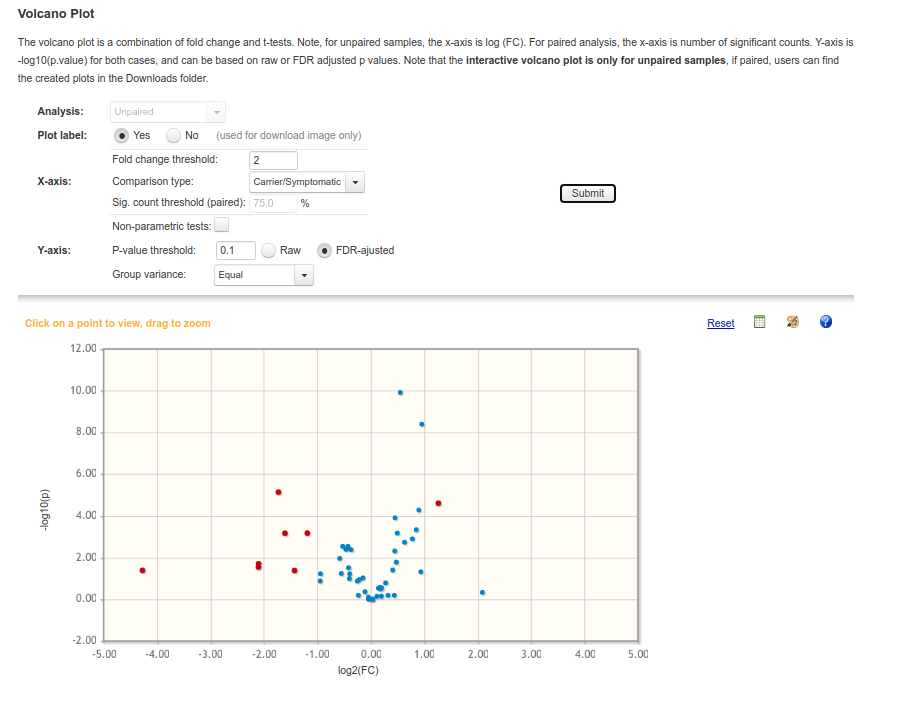


Answer: I would choose “Auto scaling” or “Range scaling”. Because by using visual inspection, we can see that the graphical summary after normalization looks like Gaussian distribution.

1. Generate an FDR-adjusted volcano plot . What Fold-change threshold did you choose and why did you choose that threshold? What P-value threshold did you choose and why did you choose that threshold? Did you assume equal or unequal group variance and why did make that choice?

The fold-change threshold I chose was 2 and p-value threshold was 0.1.

I assumed equal group variance.

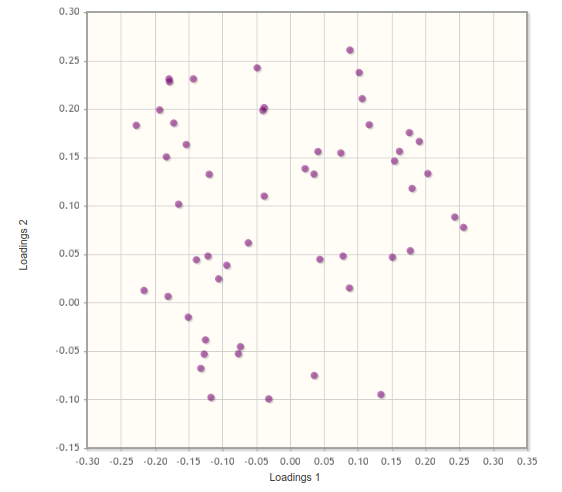


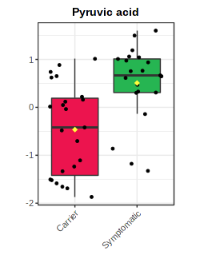
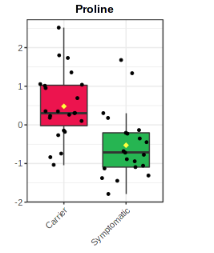
1. Perform a PCA, PLSDA, sPLSDA, and orthPLSDA on your data. Are there specific that features discriminate between asymptomatic and symptomatic carriers?

PCA score plot:



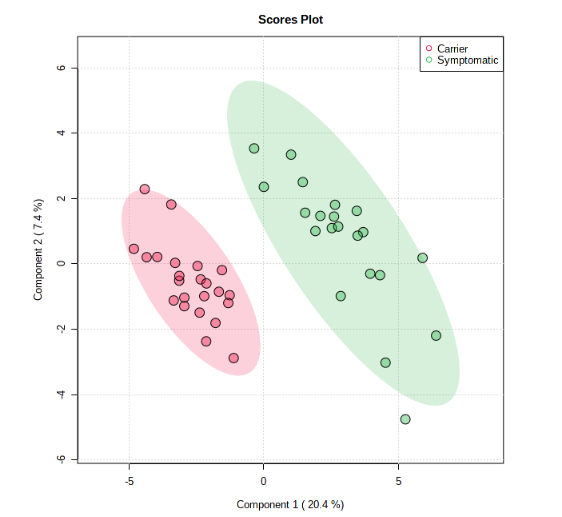
PCA loading plot:



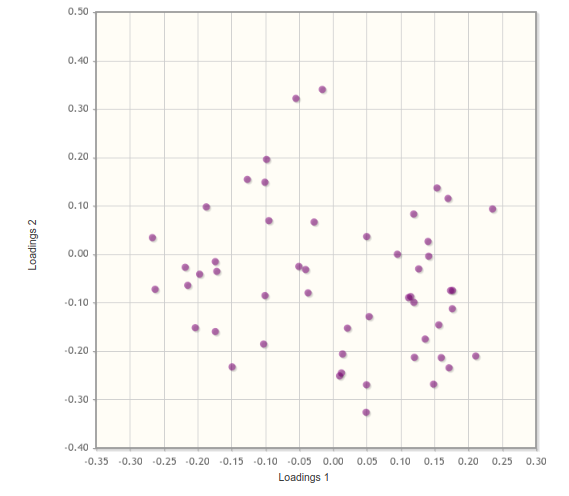


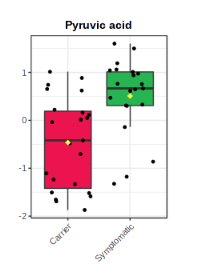
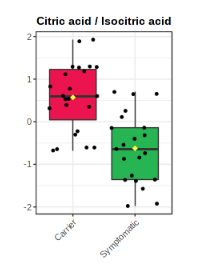
By inspecting the metabolite at the upper right and bottom left diagonal line, we can see that the proline concentration is higher in the carrier, while the pyruvic acid is higher in the symptomatic patients.

PLS-DA score plot:

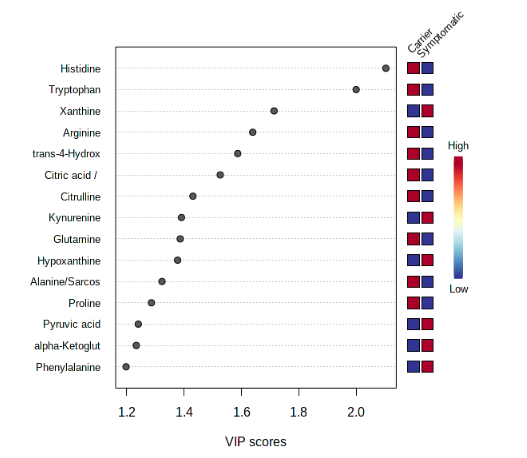


PLS-DA loading plot:



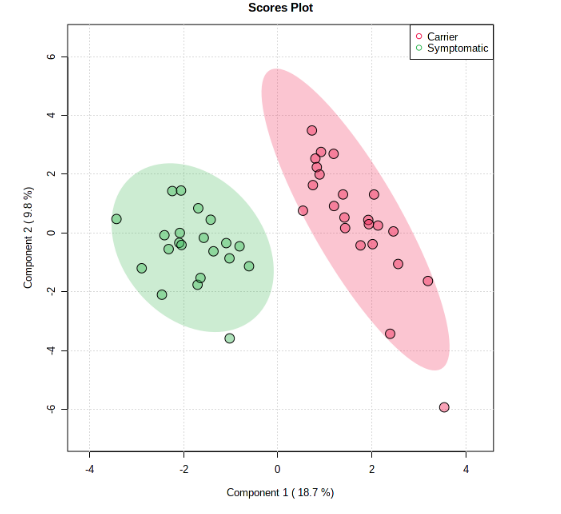


By inspecting the metabolite at the upper right and bottom left diagonal line, we can see that the citric acid/isocitric acid concentration is higher in the carrier, while the pyruvic acid is higher in the symptomatic patients.

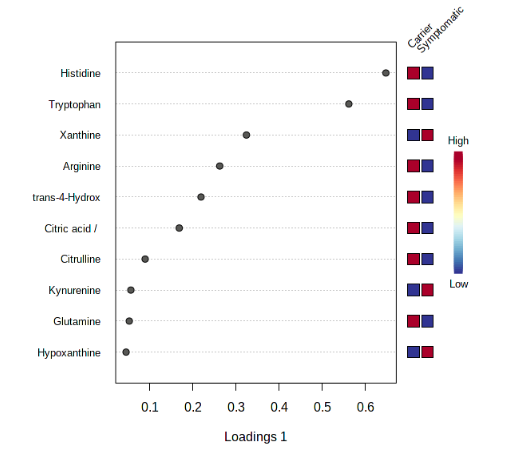


By looking at the VIP score plot, we can see that histidine has the highest score and is high in carrier.

sPLS-DA score plot:

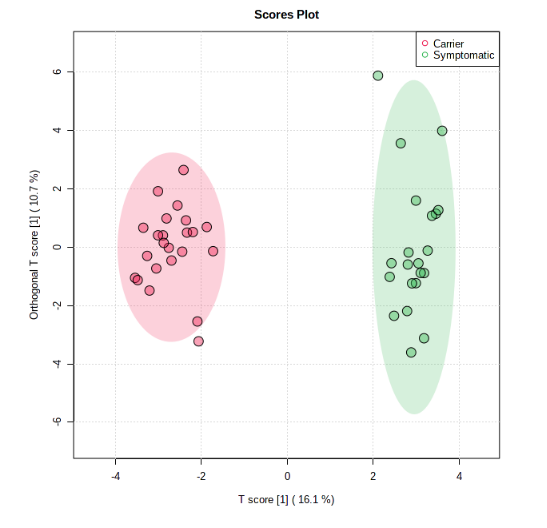


sPLS-DA loading plot:

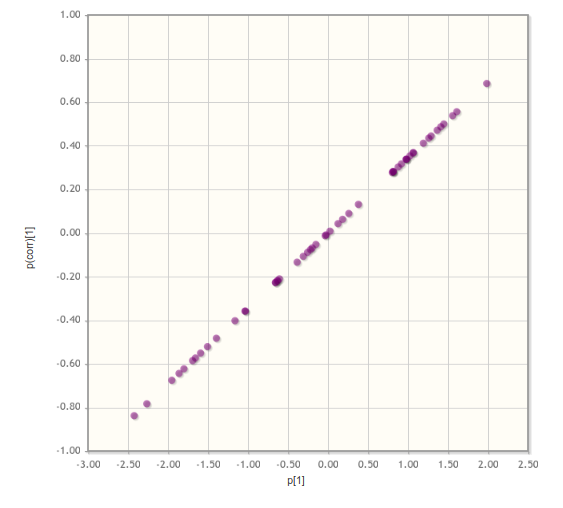


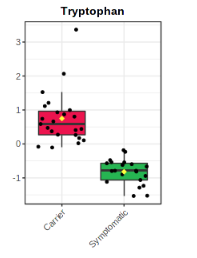
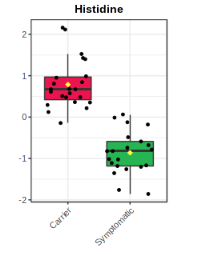
By looking at the loading 1 plot, we can see that histidine has the highest score and is high in carrier.

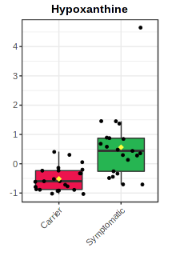
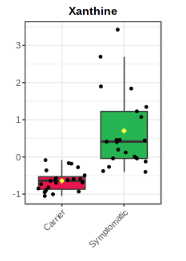
orthPLS-DA score plot:



Sig. Feature





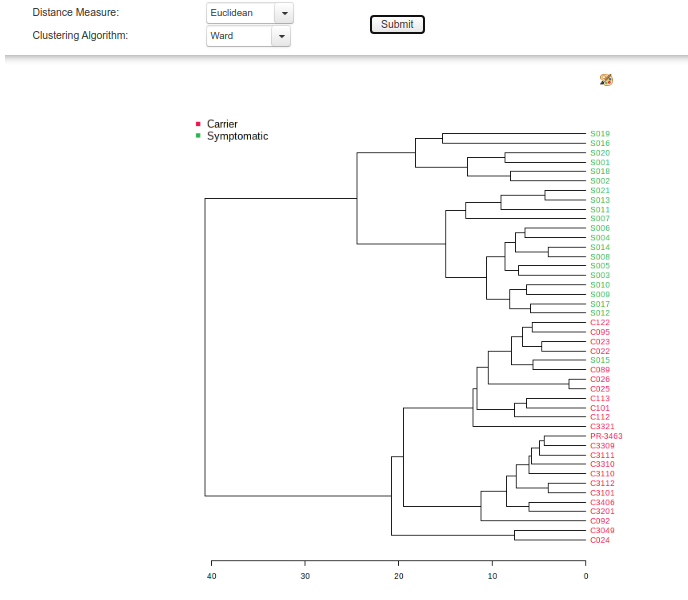


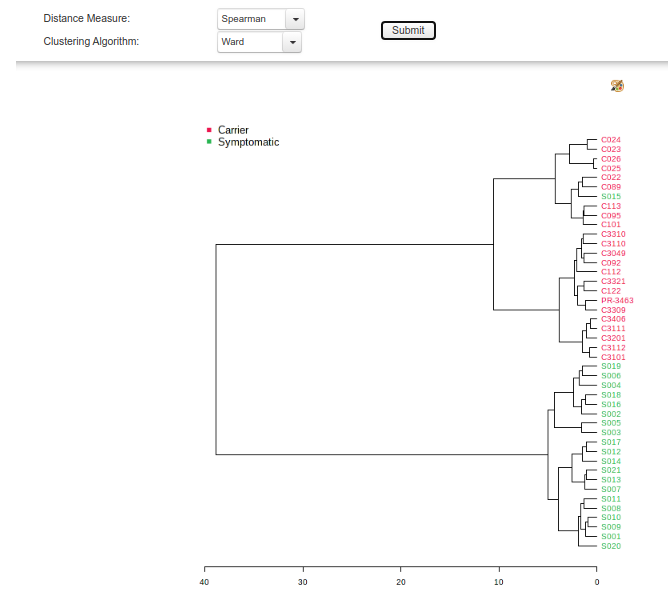
By inspecting the features at the upper right and bottom left diagonal line, we can see that the histidine and tryptophan concentration are higher in the carrier, while the xanthine and hypocanthine are higher in the symptomatic patients.

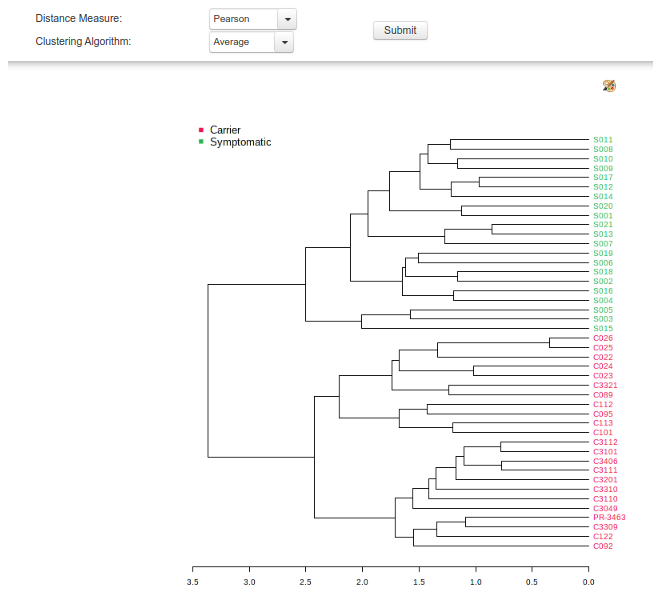
Overall, histidine, tryptopahn and xanthine might be features that discriminant between asymptomatic and symptomatic carriers.

1. Cluster your data. What does this tell you? Why did you choose the distance measure and clustering algorithm you decided to go with?

Hierarchical Clustering Dendrogram







By visual inspection, we can see that the Pearson correlation and “Average” clustering algorithm separate the carrier and symptomatic patients nicely, which minimize the distance within the groups while maximizing the distance between the groups..

**References**

Clark LN, Kisselev S, Park N, Ross B, Verbitsky M, Rios E, Alcalay RN, Lee JH, Louis ED (2010) Mutations in the Parkinson's disease genes, Leucine Rich Repeat Kinase 2 (LRRK2) and Glucocerebrosidase (GBA), are not associated with essential tremor. Parkinsonism Relat Disord 16:132-135.

Tsuang D et al. (2012) GBA mutations increase risk for Lewy body disease with and without Alzheimer disease pathology. Neurology 79:1944-1950.