**51537 – Bailey GC-MS Metabolomics Report**   
Updated: 11/03/2021

EMSL Team

GC-MS Data Processing and Analysis: Chaevien Clendinen  
Statistics:  Chaevien Clendinen  
Report and Figures: Chaevien Clendinen

Experimental

1. GC-MS Sample Derivatization and Data Acquisition

1. Dried metabolite extracts from samples are derivatized using a modified version of the protocol used to create FeihnLib.1
2. Samples undergo methoximation to protect carbonyl groups and reduce tautomeric isomers
3. Followed by silylation with N-methyl-N-trimethylsilyltrifluoroacetamide and 1% trimethylchlorosilane (MSTFA) to derivatize hydroxy and amine groups to trimethylsilated (TMS) forms.
4. Samples put into an autosampler tray for GC/MS analysis.
5. An Agilent GC 7890A coupled with a single quadrupole MSD 5975C (Agilent Technologies) is used for collection of GC/MS data.
   1. Data is collected over a mass range of 50-550 m/z.
   2. standard mixture of fatty acid methyl esters (FAMEs) (C8-C28) are analyzed with samples for RI alignment.
   3. The GC oven is held at 60 ºC for 1 min after injection followed by a temperature increase by 10 ºC min-1 to a maximum of 325 ºC at which point it will be held for 5 min.

2. GC-MS Data Processing and Analysis

1. GC-MS raw data files were processed using Metabolite Detector software, version 2.5 beta.2
2. Agilent .D files were converted to netCDF format using Agilent Chemistation.
3. Files were then converted to binary files using Metabolite Detector.
4. Retention indices (RIs) of detected metabolites were calculated based on analysis of the Fatty acid Methyl Esters standard mixture followed by chromatographic deconvolution and alignment.
5. Metabolites were initially identified by matching experimental spectra to an augmented version of FiehnLib.3
6. All metabolite identifications were manually validated with the NIST 14 GC–MS library.
7. The summed abundances of the three most abundant fragment ions of each identified metabolite were were integrated across the GC elution profile (automatically determined by Metabolite Detector); fragment ions due to trimethylsilylation (that is, m/z 73 and 147) were excluded from the determination of metabolite abundance.
8. Features resulting from GC column bleeding were removed from the data before further data processing and analysis.

3. Statistical Analysis

1. Processed GC MS data was further analyzed in MATLAB 2020b. All 0’s were replaced with NaN, log10 transformed and a global normalization to the median for statistical analysis.
2. Probabilistic Principal Components Analysis.
3. Spearman correlation between samples was calculated.

Results

1. Attached Files

Bailey\_GCMS\_Metab112021.xlsx

2 – Results

**Part 1: All Data**

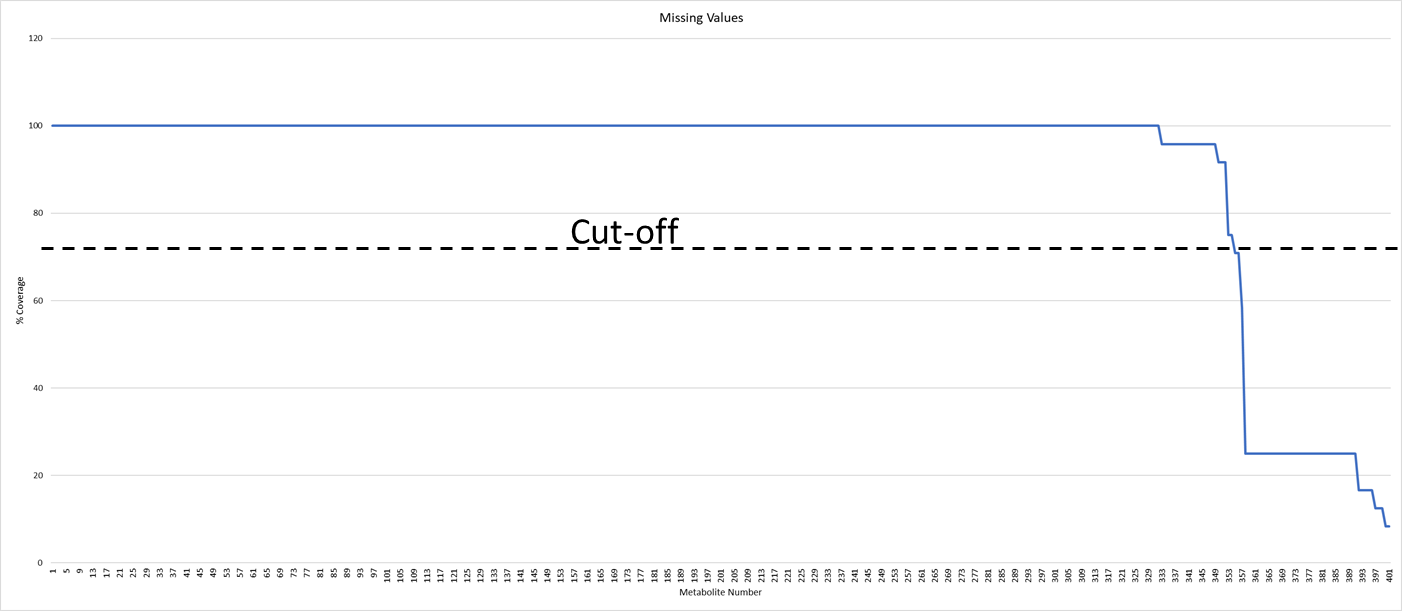


Figure 1: Metabolites with less than 70% coverage were removed (~43 features).

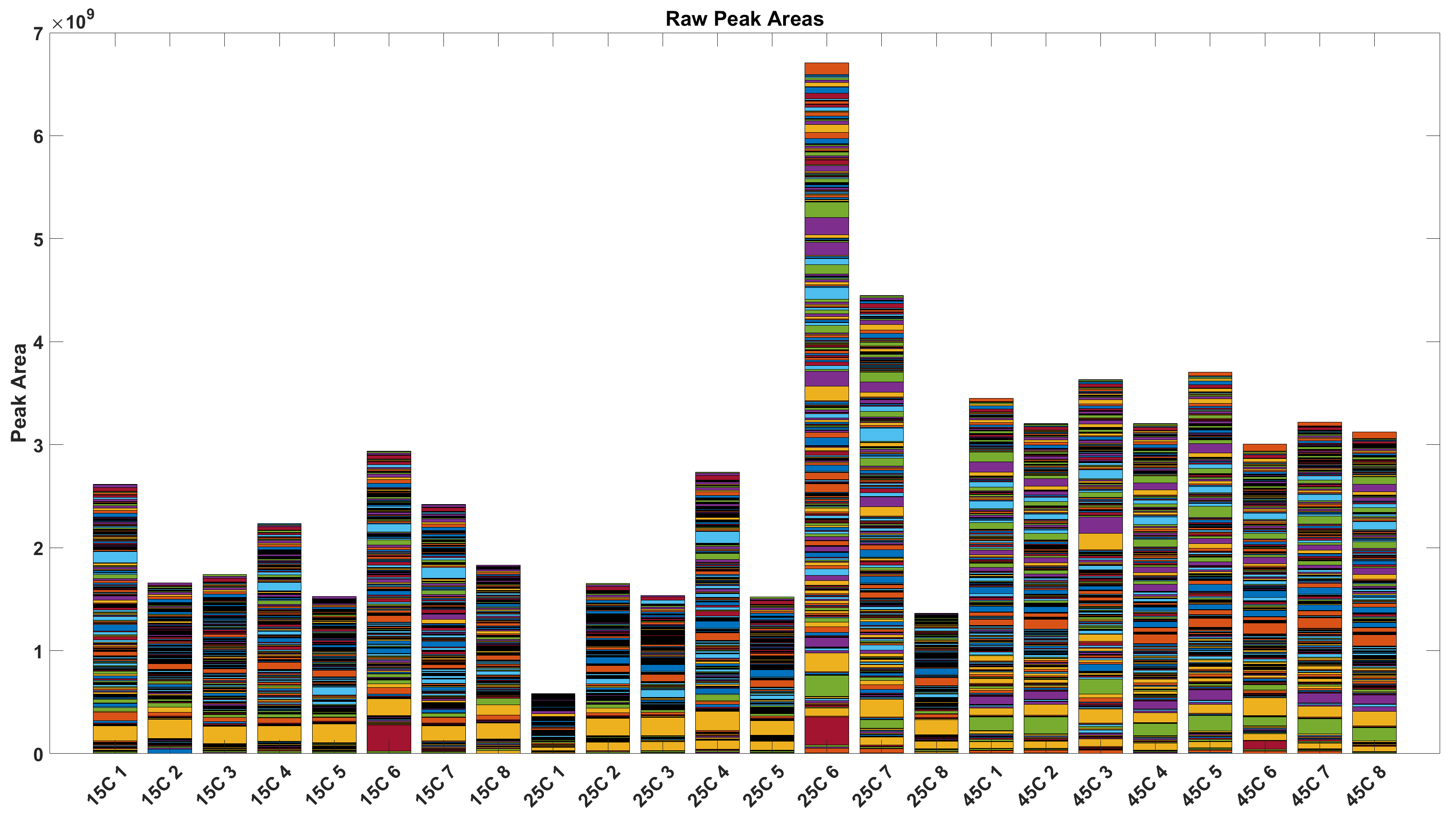


Figure 2: Total Raw peak areas.

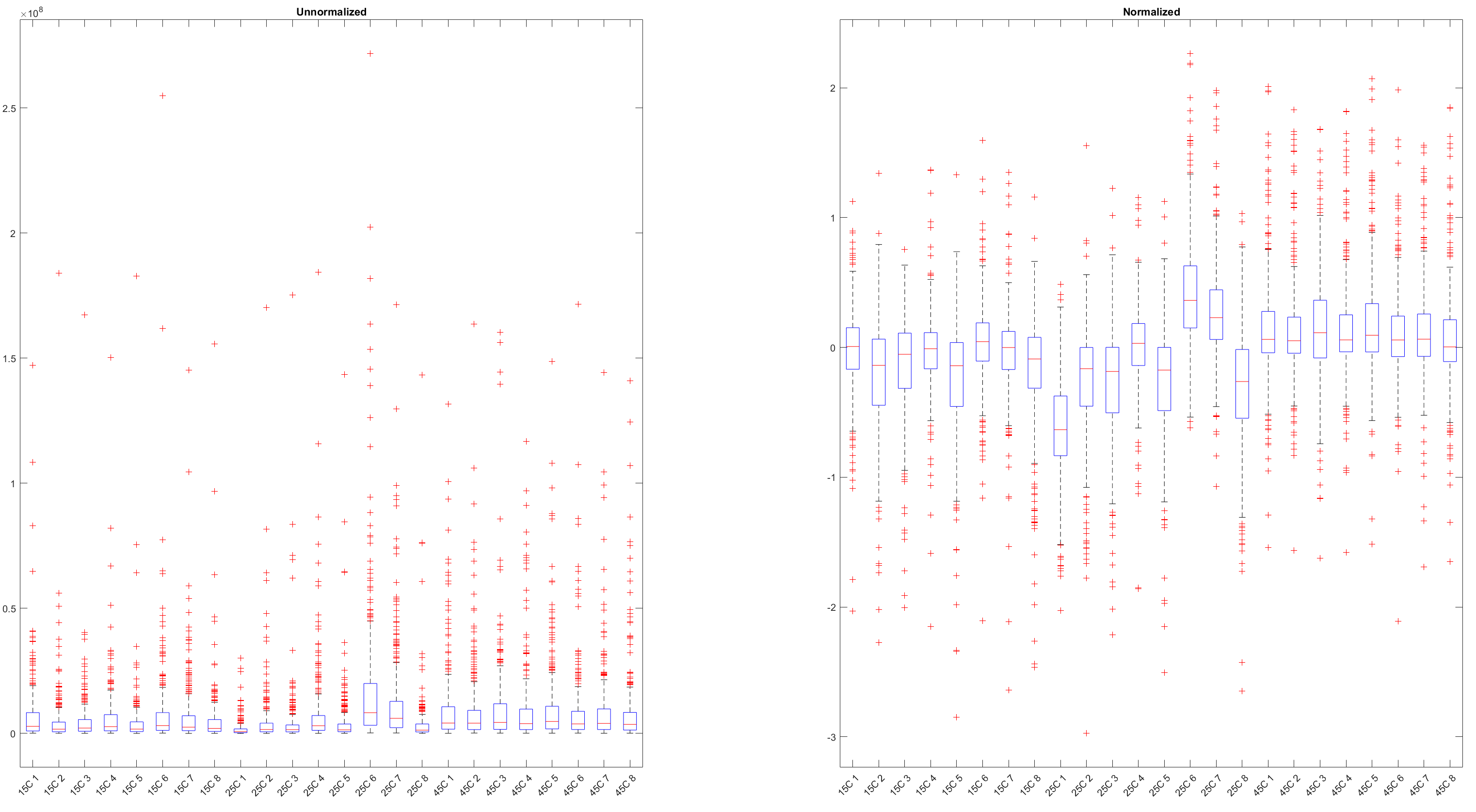


Figure 3: Unormalized and normalized data.

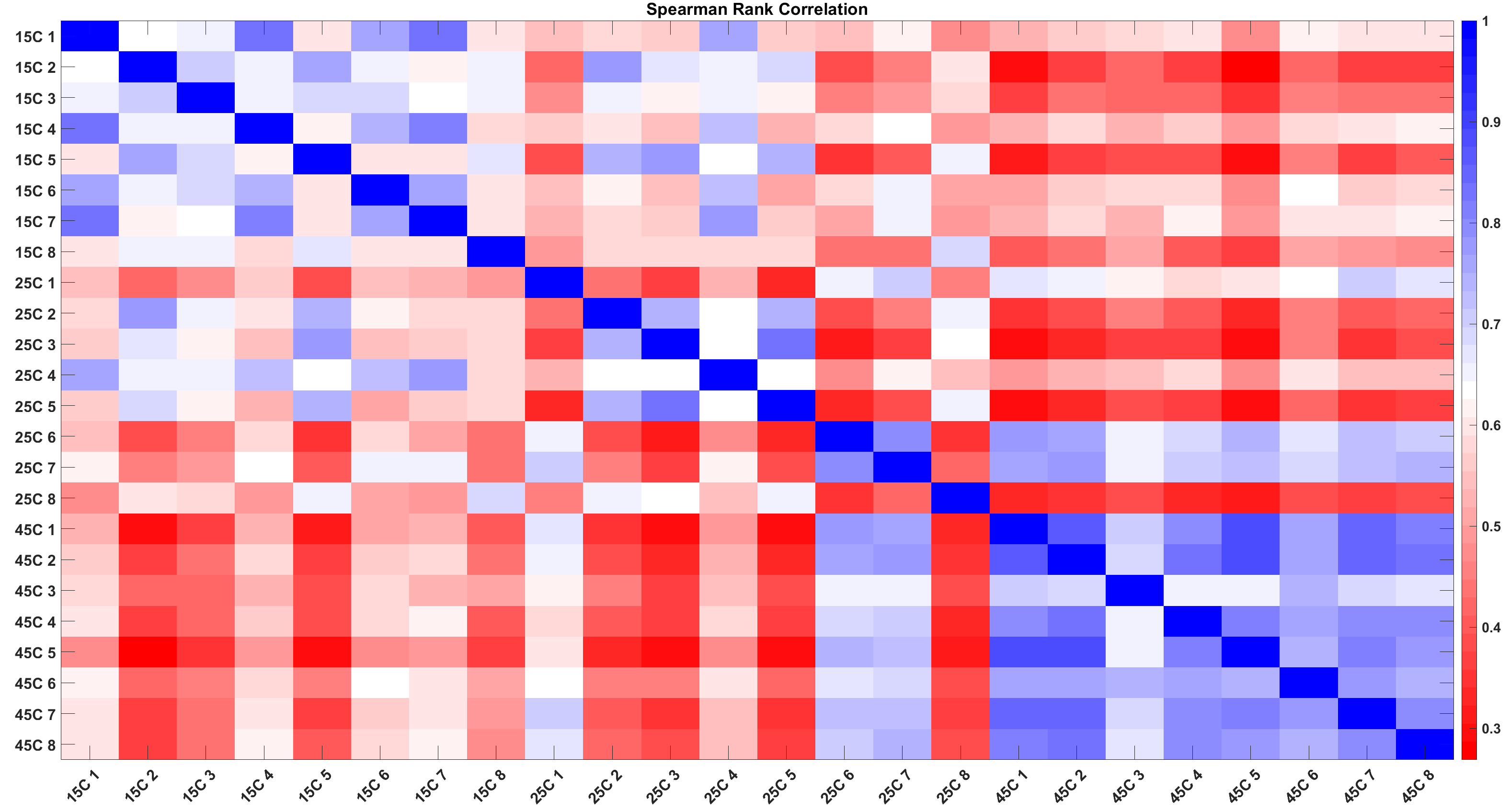


Figure 4: Spearman rank correlations on raw data.

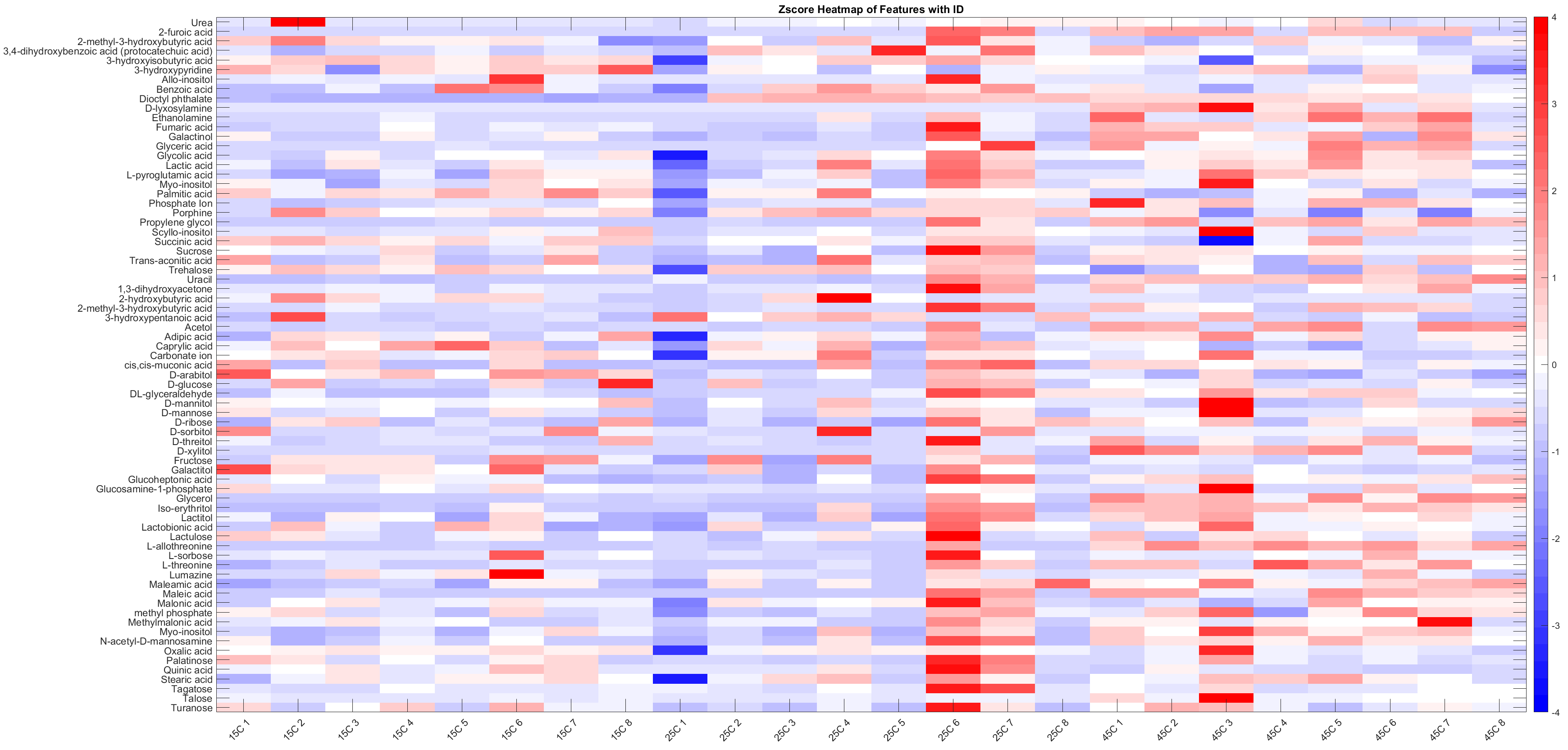


Figure 5: Zscore heatmap of assigned features

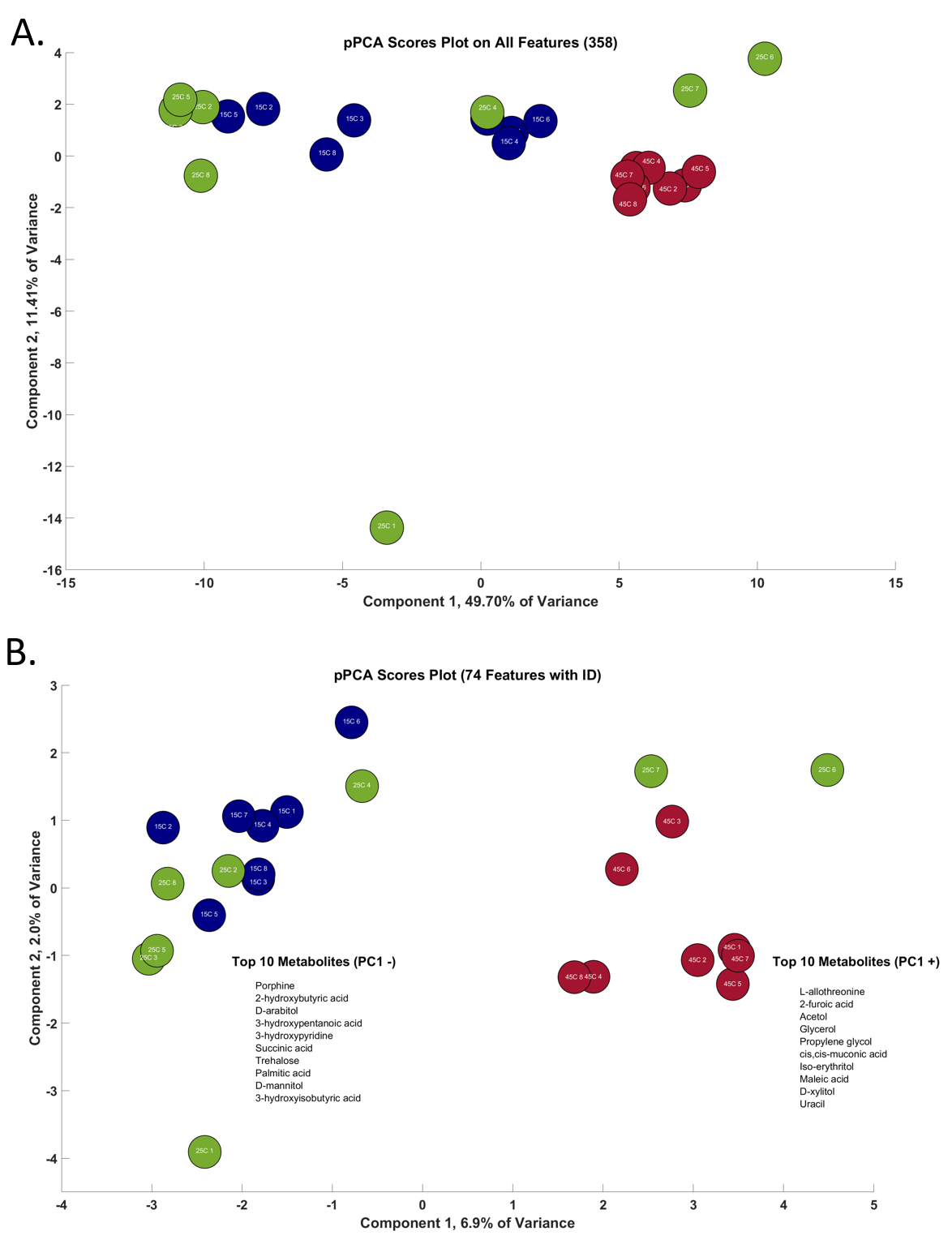


Figure 6: Probabilistic principal component analysis on all data (401) (A) and tentatively assigned (86) (B). 15C (Blue), 25C (Green), 45C (orange). Not listed in order of importance

Significance Testing (one-way ANOVA) shows that there is a significant difference overall between 15C and 45C and 25C and 45C, but no significant difference between of 15C and 25C.

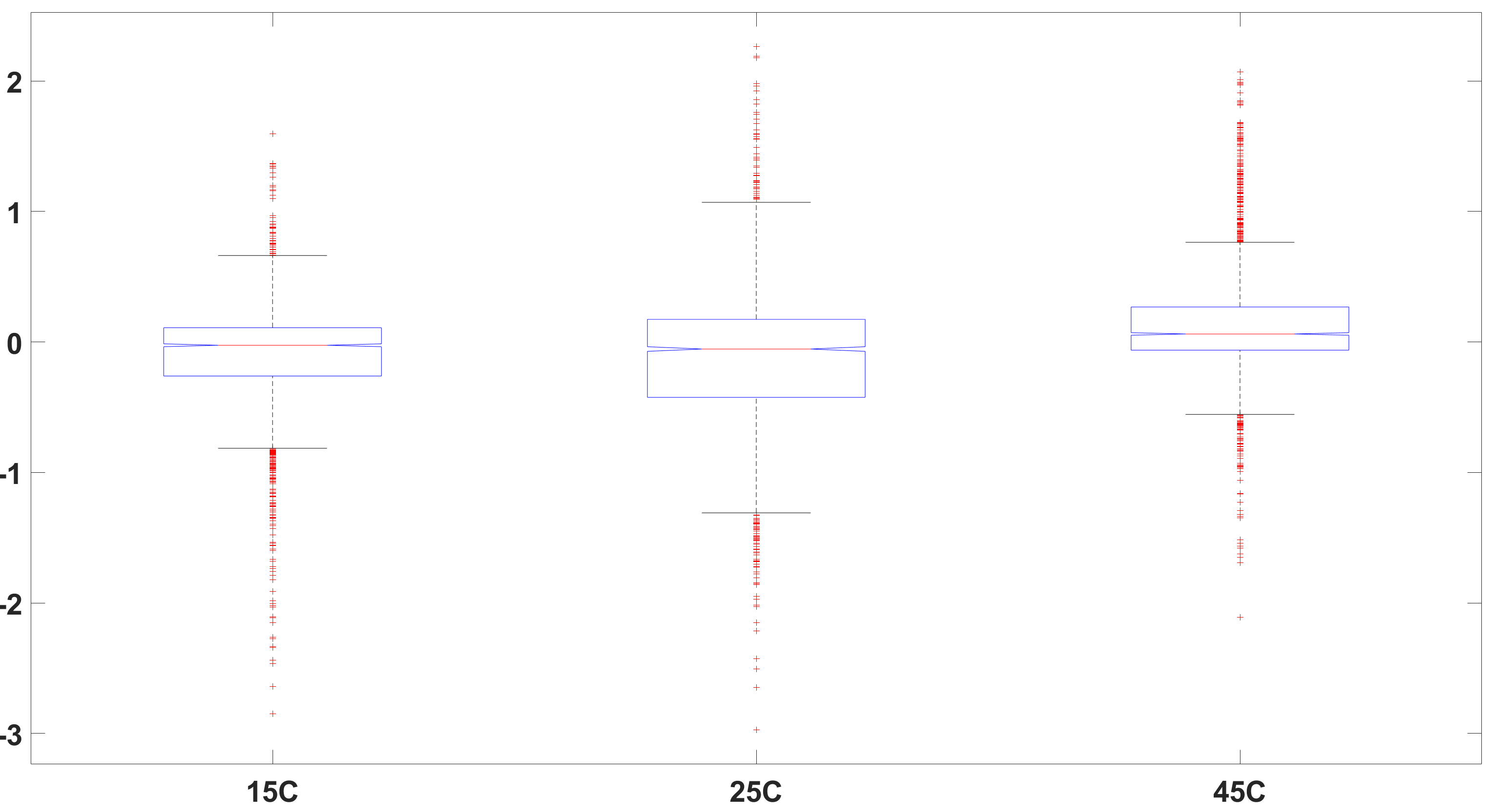


Figure 7: Box plot of the observations for each group. The central mark is the median (Q2) and the edges are the 25th and 75th percentiles (Q1 and Q3 respectively). Outliers are ploted using +. Two medians are significantly different at the 5% significance level if their intervals, represented by notches, do not overlap. Large differences in the center lines of the boxes correspond to a large F-statistic value and correspondingly a small p-value.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Source | SS | df | MS | F | Prob>F |
| Groups | 109.68 | 2 | 54.84 | 262.68 | 2.053e-111 |
| Error | 1772.05 | 8488 | 0.2088 |  |  |
| Total | 1881.73 | 8490 |  |  |  |

References

1. Fiehn, O., Metabolomics by Gas Chromatography-Mass Spectrometry: Combined Targeted and Untargeted Profiling. *Curr Protoc Mol Biol* **2016,** *114*, 30 4 1-30 4 32.

2. Hiller, K.; Hangebrauk, J.; Jager, C.; Spura, J.; Schreiber, K.; Schomburg, D., MetaboliteDetector: comprehensive analysis tool for targeted and nontargeted GC/MS based metabolome analysis. *Anal Chem* **2009,** *81* (9), 3429-39.

3. Kind, T.; Wohlgemuth, G.; Lee, D. Y.; Lu, Y.; Palazoglu, M.; Shahbaz, S.; Fiehn, O., FiehnLib: mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. *Anal Chem* **2009,** *81* (24), 10038-48.