

R for multi-omics analysis

uni-variate, multi-variate and integration

Shuang PENG

AgroParisTech

2022-11-21

Multi-omics Data types

Proteomic table

| sample name | temperature | AT1G01470 | ... | AT1G02700 |
|-------------|-------------|------------|-----|------------|
| low r1 | Low | 22866363 | ... | 48307337.4 |
| low r2 | Low | 19209302.9 | ... | 48089559 |
| low r3 | Low | 24780232.3 | ... | 70132972.3 |
| medium r1 | Medium | 46092725 | ... | 79060256 |
| medium r2 | Medium | 49154335.8 | ... | 85218205.4 |
| medium r3 | Medium | 36031052.7 | ... | 82237069.8 |
| elevated r1 | Elevated | 47252917.7 | ... | 97170910 |
| elevated r2 | Elevated | 29018829.6 | ... | 48648217.4 |
| elevated r3 | Elevated | 33876783.2 | ... | 70496404.5 |

Metabolomic table

| sample name | temperature | Fumarate | ... | Sulfite |
|-------------|-------------|-------------|-----|-------------|
| low r1 | Low | 45580432 | ... | 192791869.9 |
| low r2 | Low | 99058820.5 | ... | 144595795 |
| low r3 | Low | 108103566.8 | ... | 104077985.5 |
| medium r1 | Medium | 42064517.3 | ... | 172892634 |
| medium r2 | Medium | 60077401.4 | ... | 150221343 |
| medium r3 | Medium | 68995980.3 | ... | 207105263.2 |
| elevated r1 | Elevated | 71228127.9 | ... | 247807516.5 |
| elevated r2 | Elevated | 84058902.7 | ... | 238145287.4 |
| elevated r3 | Elevated | 88945490.5 | ... | 241025090.6 |

multi-omic data table

Analysis workflow

Proteomic table

| sample name | temperature | AT1G01470 | ... | AT1G02700 |
|-------------|-------------|------------|-----|------------|
| low r1 | Low | 22866363 | ... | 48307337.4 |
| low r2 | Low | 19209302.9 | ... | 48089559 |
| low r3 | Low | 24780232.3 | ... | 70132972.3 |
| medium r1 | Medium | 46092725 | ... | 79060256 |
| medium r2 | Medium | 49154335.8 | ... | 85218205.4 |
| medium r3 | Medium | 36031052.7 | ... | 82237069.8 |
| elevated r1 | Elevated | 47252917.7 | ... | 97170910 |
| elevated r2 | Elevated | 29018829.6 | ... | 48648217.4 |
| elevated r3 | Elevated | 33876783.2 | ... | 70496404.5 |

Metabolomic table

| sample name | temperature | Fumarate | ... | Sulfite |
|-------------|-------------|-------------|-----|-------------|
| low r1 | Low | 45580432 | ... | 192791869.9 |
| low r2 | Low | 99058820.5 | ... | 144595795 |
| low r3 | Low | 108103566.8 | ... | 104077985.5 |
| medium r1 | Medium | 42064517.3 | ... | 172892634 |
| medium r2 | Medium | 60077401.4 | ... | 150221343 |
| medium r3 | Medium | 68995980.3 | ... | 207105263.2 |
| elevated r1 | Elevated | 71228127.9 | ... | 247807516.5 |
| elevated r2 | Elevated | 84058902.7 | ... | 238145287.4 |
| elevated r3 | Elevated | 88945490.5 | ... | 241025090.6 |

single-omic analysis

by single omic data

uni-variate analysis

approach

data exploration

differentiate test

multi-variate analysis

approach

PCA

hierarchical clustering

multi-omic integration

by all omic data

data exploration

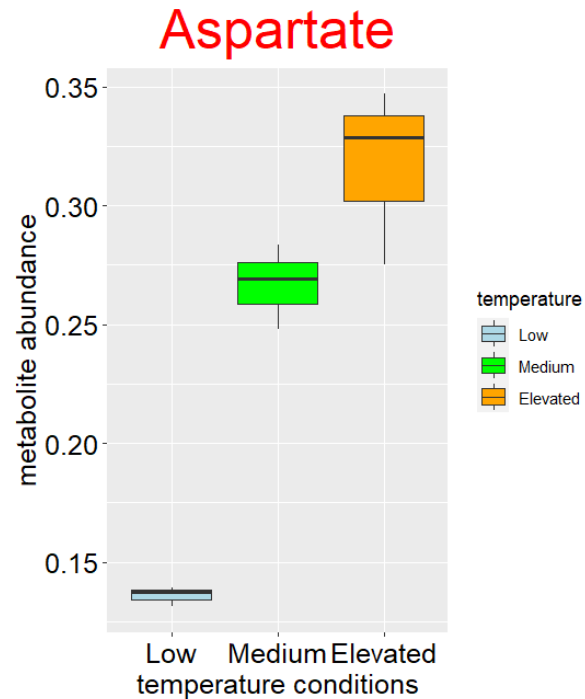
co-expression analysis

Uni-variate analysis

BoxPlot

One-Factor ANOVA

Multi-Comparison



boxplot of asparate

```
# remplissage des box selon modalités de temperature
ggplot(table2, aes(fill=temperature)) +
  geom_boxplot(aes_string(x=temperature, y=table2[,i])) +

# formatage
labs(
  title=varname,
  x = "temperature conditions",
  y = "metabolites abundance"
) +
theme(
  plot.title = element_text(colour="red", size=30,hjust=0.5),
  axis.title.x = element_text(colour="black", size=15),
  axis.title.y = element_text(colour="black", size=15),
  axis.text.x = element_text(colour="black", size=15),
  axis.text.y = element_text(colour="black", size=15)
) +
scale_fill_manual(
  breaks = c("Low", "Medium", "Elevated"),
  values=c("lightblue", "green", "orange"))
```

Uni-variate analysis

BoxPlot

One-Factor ANOVA

Multi-Comparison

```
aov(table2[,i]~temperature, data=table2)
```

Call:

```
aov(formula = table2[, 5] ~ temperature, data = table2)
```

Terms:

| | temperature | Residuals |
|-----------------|-------------|------------|
| Sum of Squares | 0.05236010 | 0.00346879 |
| Deg. of Freedom | 2 | 6 |

Residual standard error: 0.02404438

Estimated effects may be unbalanced

```
summary(aov(table2[,i]~temperature, data=table2))
```

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|-------------|----|---------|----------|---------|-------------|
| temperature | 2 | 0.05236 | 0.026180 | 45.28 | 0.00024 *** |
| Residuals | 6 | 0.00347 | 0.000578 | | |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
pvalue.anova.adj <- p.adjust(  
  pvalue.anova,  
  method="fdr"  
)
```

```
table(pvalue.anova<0.05)  
FALSE TRUE  
65 70
```

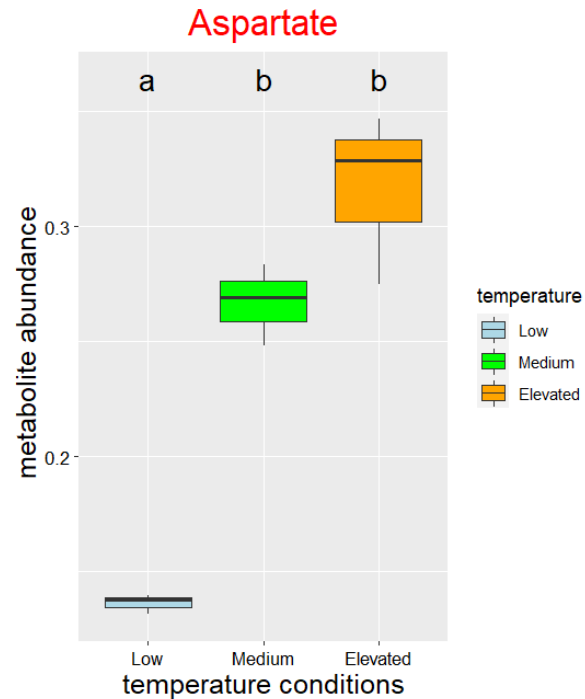
```
table(pvalue.anova.adj<0.05)  
FALSE TRUE  
73 62
```

Uni-variate analysis

BoxPlot

One-Factor ANOVA

Multi-Comparison



```
TukeyHSD(aov(table2[,i]~temperature, data=table2))
```

Tukey multiple comparisons of means
95% family-wise confidence level

```
Fit: aov(formula = table2[, 5] ~ temperature, data = table2)
```

\$temperature

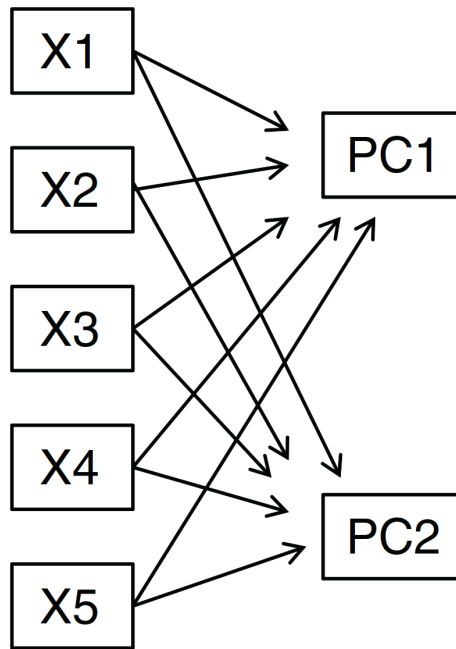
| | diff | lwr | upr | p adj |
|-----------------|------------|-------------|-----------|-----------|
| Medium-Low | 0.13091624 | 0.07067942 | 0.1911531 | 0.0013400 |
| Elevated-Low | 0.18089536 | 0.12065853 | 0.2411322 | 0.0002267 |
| Elevated-Medium | 0.04997911 | -0.01025771 | 0.1102159 | 0.0962989 |

boxplot of aspartate with Tukey
test

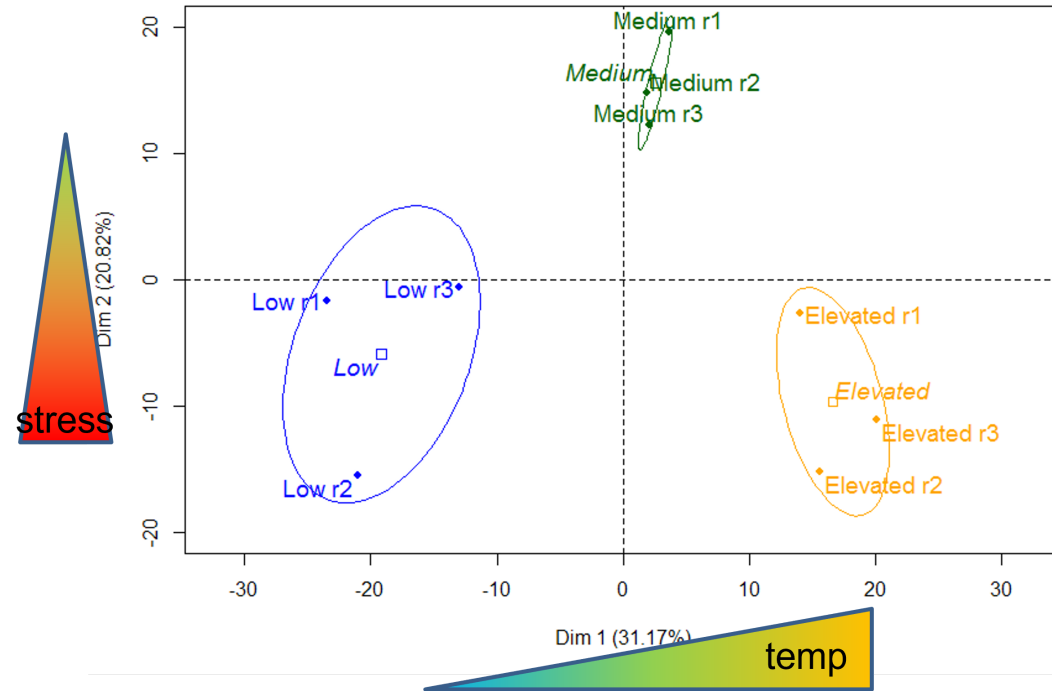
Multi-variate analysis

Principal Component Analysis (PCA)

Hierarchical Clustering



R in action (2nd edition).
Kabacoff et al. 2015.

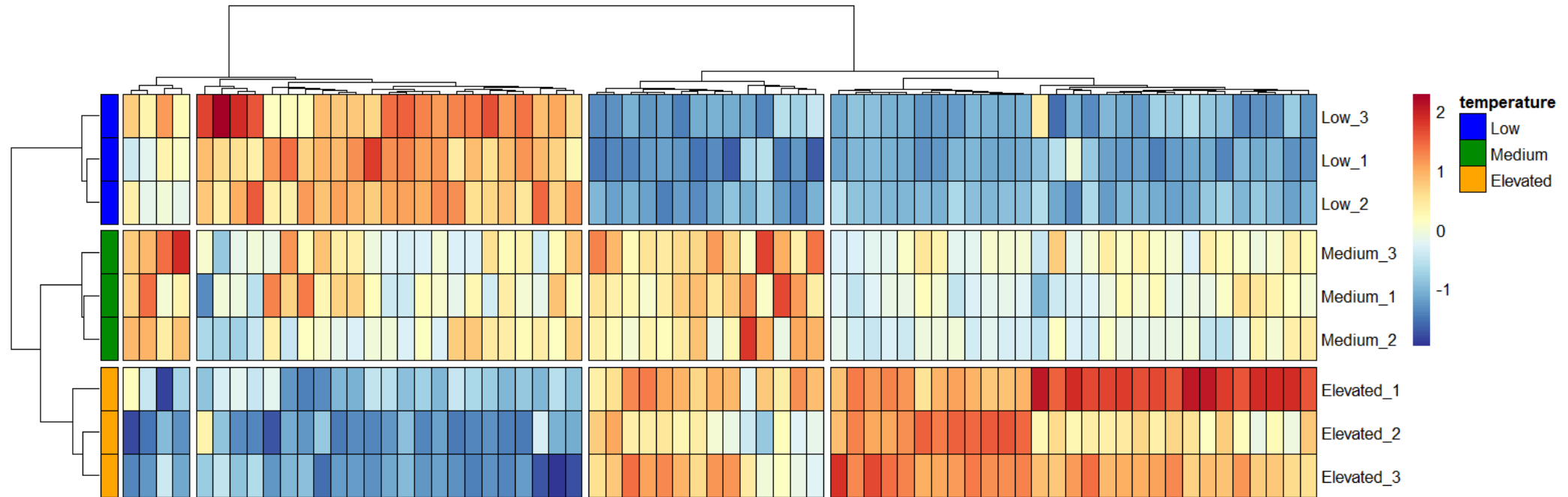


PCA projection by PC1 & PC2 (proteomic data)

Multi-variate analysis

Principal Component Analysis (PCA)

Hierarchical Clustering

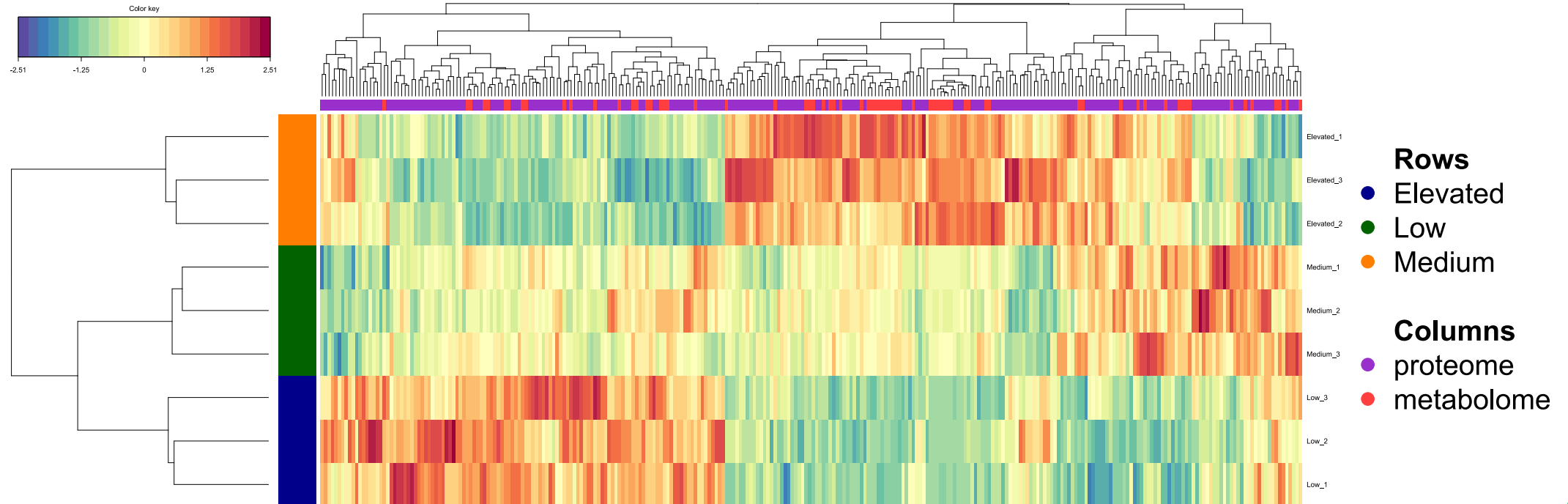


Heatmap of 70 selected metabolites

Multi-Omics integration

Integrated hierarchial clustering

co-expression analysis (mixOmics)

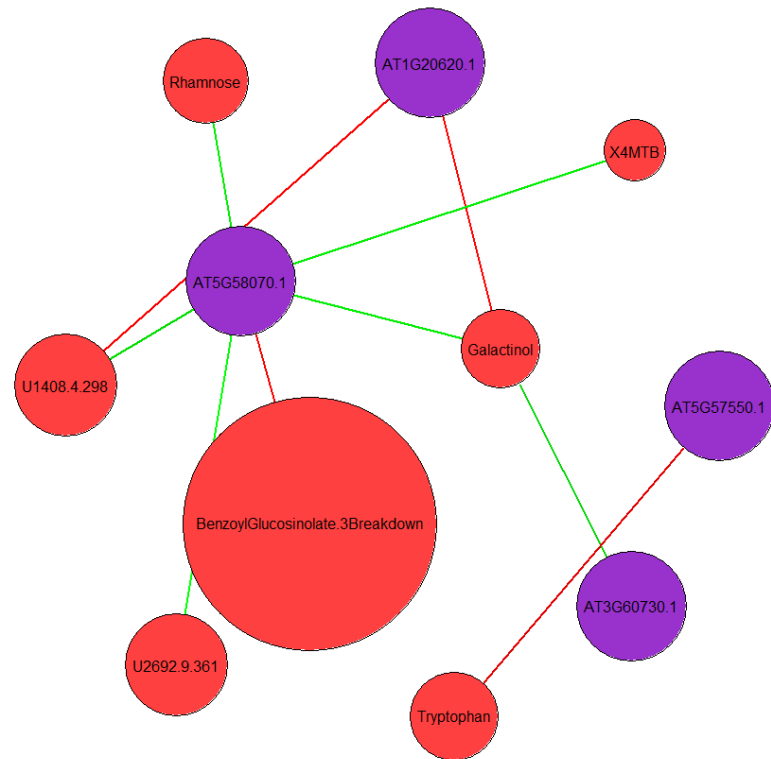


Heatmap of selected metabolites & proteins

Multi-Omics integration

Integrated hierarchial clustering

co-expression analysis (mixOmics)



network visualisation

