Vignette1 - Understanding the Results of MCIA

Max Mattessich   Joaquin Reyna   Anna Konstorum

9/1/2022

Part 1: Interpreting Global Factor Scores

The main MCIA function `nipals_multiblock` outputs a global factor score matrix $\mathbf{F}$ that is $n \times r$, where n is the number of samples and r is the number of factors chosen by the user with the `num_PCs` argument. Each column of this matrix represents one of the orders of global factors computed, i.e.

$$\mathbf{F} = \begin{pmatrix} | & | & & | \\ \mathbf{f}^{(1)} & \mathbf{f}^{(2)} & \cdots & \mathbf{f}^{(r)} \\ | & | & & | \end{pmatrix}.$$

Each entry M_{ij} represents the strength of factor j to sample i. As such, you may be interested in further dissecting the global scores matrix in order to understand, for a given sample, what factors capture the most the MCIA decomposition... (to be continued)

Running MCIA on NCI60 Data and basic visualization

```
data(NCI60)
results <- nipals_multiblock(data_blocks, preprocMethod='colprofile',num_PCs = 10, tol=1e-12)
```

```
## Performing centered column profile pre-processing...
```
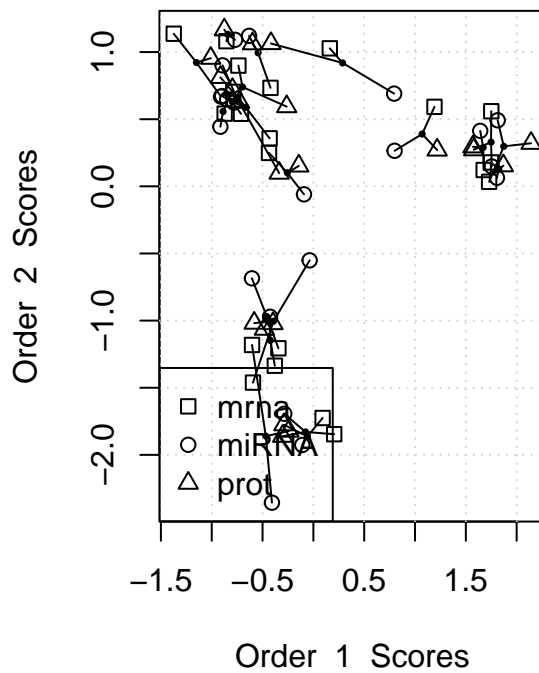
```
## Pre-processing completed.
```

```
## Computing order 1 scores
```

```
## Computing order 2 scores
```

```
## Computing order 3 scores
```

```
## Computing order 4 scores
```

```
## Computing order 5 scores
```

```
## Computing order 6 scores
```

```
## Computing order 7 scores
```

```
## Computing order 8 scores

## Computing order 9 scores

## Computing order 10 scores
```
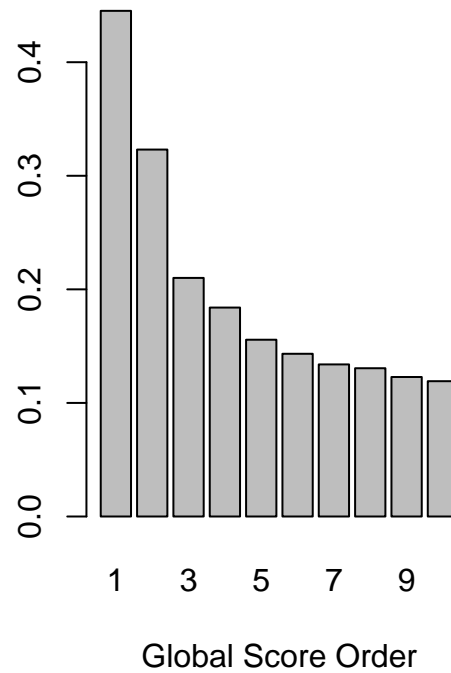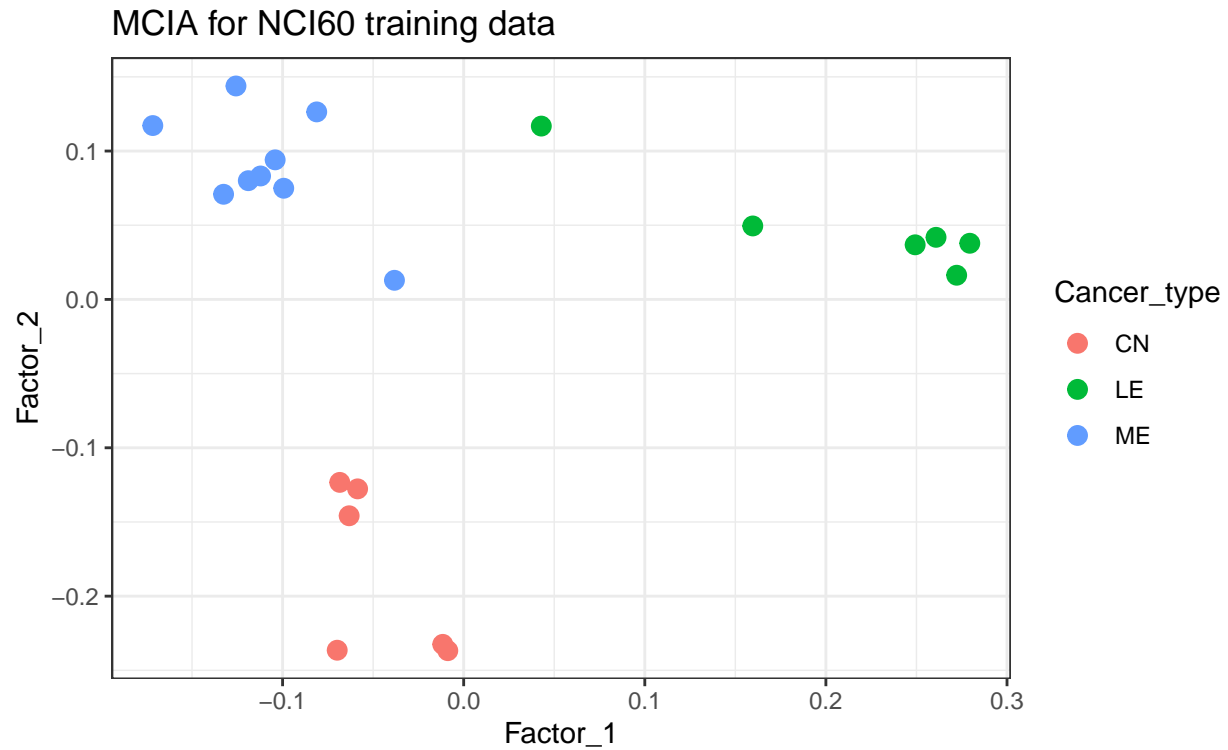

Visualize only global factor scores

```r
cancer_type <- substr(rownames(data_blocks$mrna), 1, 2)
MCIA_out<-data.frame(results$global_scores[,1:2])
MCIA_out$Cancer_type<-cancer_type
colnames(MCIA_out)<-c("Factor_1", "Factor_2", "Cancer_type")
ggplot(data = MCIA_out, aes(x=Factor_1, y=Factor_2, color=Cancer_type))+
  geom_point(size=3) +
  theme_bw() +
  ggtitle("MCIA for NCI60 training data")
```

MCIA for NCI60 training data

Visualizing the clustering of samples by factor scores

```r
gs_scores = results$global_scores
colnames(gs_scores) = paste0('F', seq(1, ncol(results$global_scores)))
p = ComplexHeatmap::Heatmap(gs_scores,
    name = "GS Score",
    column_title = "Factors",
    row_title = "Samples",
    row_names_gp = grid::gpar(fontsize = 7),
    show_column_names = T,
    show_row_names = T,
    row_names_side = "right"
)
p
```
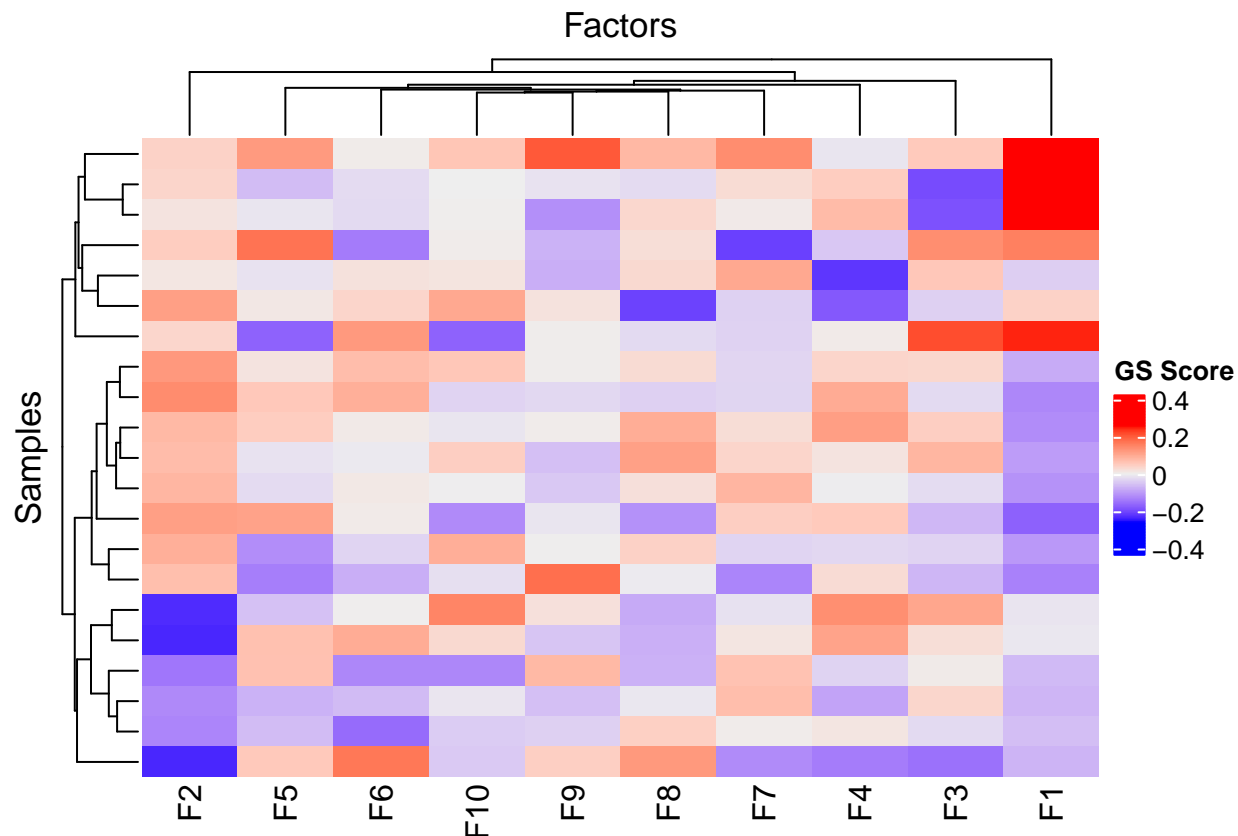
Part 2: Interpreting Global Loadings

In addition to the global scores matrix, MCIA also calculates a global loadings matrix that is $(m_1 + ... + m_j + ... + m_R) \times k$ where m_j is the number of features within the omics matrix X^j and K is the number of factors calculated. This second matrix provides information as to the contribution

Pseudoeigenvalues representing the contribution of each omic to the global factor score

. . . (to be continued)

Scree Plot: Visualizing the top features per factor

. . . (to be continued)

Pathway analysis for the top factors using data from gene-centric omics blocks

. . . (to be continued)