Omics data integration with the O2PLS R-package

Said el Bouhaddani, Jeanine Houwing-Duistermaat, Hae-Won Uh

April 4, 2016

O2PLS Method

- Trygg & Wold, 2003
- Decomposition:

$$X = TW^\top + T_\perp W_\perp^\top + E$$

$$Y = UC^\top + U_\perp C_\perp^\top + F$$

- Joint part: $U = TB + H$
- Find W and C such that T and U have high covariance.
- W and C corrected for independent latent variation specific for X and Y.
- n joint components, nx X-specific components, ny Y-specific components

O2PLS R package: Overview

- Input data `X` and `Y`, rows are **the same** subjects
- Number of components `n`, `nx` and `ny`.
  - Main fitting function `o2m(X, Y, n, nx, ny)`
  - *Simultaneous* estimation of all components per part
  - Stripped version is also present: `stripped = TRUE`
  - Automatic switching to high dimensional mode with `p_thresh = 3000`
  - Output: list of class `o2m`
- `print/plot/summary/predict/loadings`: see `help("___.o2m", "O2PLS")`
- For a complete overview: `?O2PLS`

Data to be analyzed

- DILGOM population study
- Gene expression ($p = 6272$)
- Metabolites ($q = 137$)
- $N = 191$ participants

Data analysis

- Install from Github with devtools:
- `devtools::install_github("selbouhaddani/O2PLS", ref = "Upgrade")`
- `library(O2PLS)`
- Fit O2PLS: X = RNA, Y = Metabolites
- Low dimensional mode, since $q < 3000$.

```
fit <- o2m(X, Y, n = 1, nx = 8, ny = 1)
```

- *Number of components chosen with cross-validation, see paper*

Inspecting the results

```
fit
```

```
> O2PLS fit
> with 1 joint components
> and  8 orthogonal components in X
> and  1 orthogonal components in Y
> Elapsed time: 1.41 sec
```

Inspecting the results

```
fit
```

```
> O2PLS fit
> with 1 joint components
> and  8 orthogonal components in X
> and  1 orthogonal components in Y
> Elapsed time: 1.41 sec
```

Some timings (i5 laptop) with `stripped = TRUE`

Problem size	Timing Low D	Timing High D
1000 vars	3 sec	15 sec
5000 vars	300 sec	130 sec

Summarizing the results

```
summary(fit)
```

```
*** Summary of the O2PLS fit ***
-  Call: o2m(X = X, Y = Y, n = 1, nx = 8, ny = 1)

-  Modeled variation
-- Total variation:
in X: 116016.8
in Y: 2516.821

-- Joint, Orthogonal and Noise as proportions:

           data X data Y
Joint       0.013  0.522
Orthogonal  0.490  0.069
Noise       0.497  0.410
[TRUNCATED...]
```

Plotting the loadings

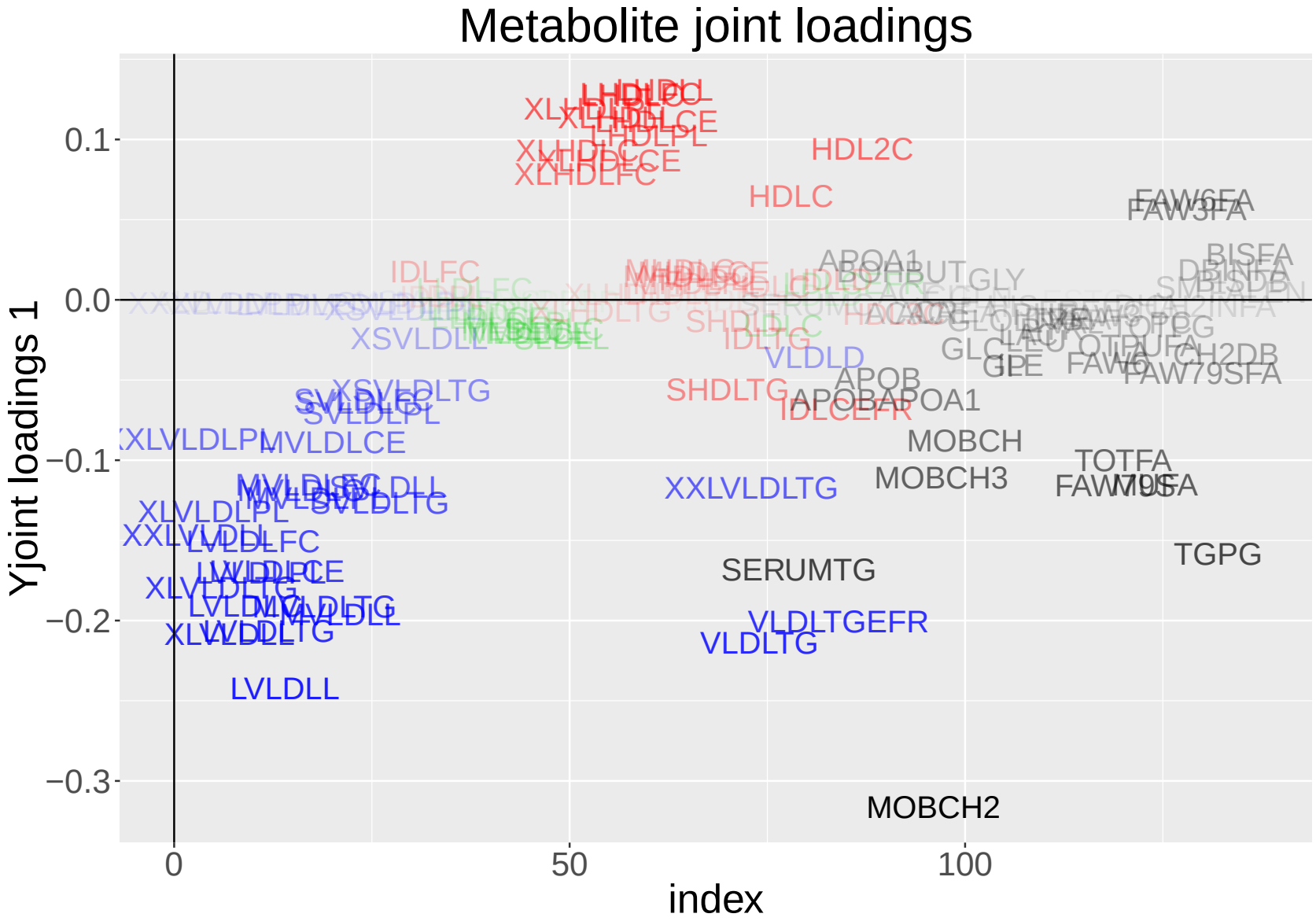
- Plot the loadings: `plot(fit, loading_name, i, j, use_ggplot2, label, ...)`
- Returns ggplot2 object
- Example: some fancy extra's:

```
alp <- loadings(fit, "Yjoint", 1) %>% abs %>% sqrt
```

- `cols` contains labeling {VLDL, LDL, HDL, other}

```
plot(fit, "Yj", i=1, label = "colnames",
     size = 6, alpha = alp/max(alp), col = cols) +
  theme(text = element_text(size = 22)) +
  ggtitle("Metabolite joint loadings")
```

Previous code results in:



Summary

- O2PLS package for omics data analysis
- Install via Github: `selbouhaddani/O2PLS`, branch `Upgrade`
- Overview of Package: `?O2PLS`
- Main function: `o2m`, see also `?o2m`

Future work

- Release on CRAN + Vignette!
- Upgrade (automated and parallelized) cross-validation
- Enhancing plots of loadings/scores

Remarks

- Acknowledgments
  - Geurt Jongbloed, TU Delft
  - Szymon Kielbasa, LUMC
- Try it out in your research!
- Please cite if you use it:

```
citation("O2PLS")
```

```
el Bouhaddani, S., Houwing-duistermaat, J., Jongbloed, G.,
Salo, P., Perola, M., & Uh, H.-W. (2016).
Evaluation of O2PLS in Omics data integration.
BMC Bioinformatics BMTL Supplement
doi:10.1186/s12859-015-0854-z
```

Finally

- The name *O2PLS* is a registered trademark
- Alternatives:
  - A. o2plsR
  - B. TwOpls
  - C. o2pls4Omics
  - D. OMICSpls

Finally

- The name *O2PLS* is a registered trademark
- Alternatives:
  - A. o2plsR
  - B. TwOpls
  - C. o2pls4Omics
  - D. OMICSpls
- Who votes
  - A. o2plsR

Finally

- The name *O2PLS* is a registered trademark
- Alternatives:
  - A. o2plsR
  - B. TwOpls
  - C. o2pls4Omics
  - D. OMICSpls
- Who votes
  - A. o2plsR
  - B. TwOpls

Finally

- The name *O2PLS* is a registered trademark
- Alternatives:
  - A. o2plsR
  - B. TwOpls
  - C. o2pls4Omics
  - D. OMICSpls
- Who votes
  - A. o2plsR
  - B. TwOpls
  - C. o2pls4Omics

Finally

- The name *O2PLS* is a registered trademark
- Alternatives:
  - A. o2plsR
  - B. TwOpls
  - C. o2pls4Omics
  - D. OMICSpls
- Who votes
  - A. o2plsR
  - B. TwOpls
  - C. o2pls4Omics
  - D. OMICSpls

Finally

- The name *O2PLS* is a registered trademark
- Alternatives:
  - A. o2plsR
  - B. TwOpls
  - C. o2pls4Omics
  - D. OMICSpls
- Who votes
  - A. o2plsR
  - B. TwOpls
  - C. o2pls4Omics
  - D. OMICSpls
- Thank you!