Multivariate Data Analysis Linking Structural and Functional Properties of High-Density Lipoproteins

Panagiotis G. Togias[1], Evangelia Zvintzou[3], George C. Sakellaropoulos[2], Kyriakos E. Kypreos[3]

[1]University of Patras Medical School, [2]Department of Medical Physics, University of Patras Medical School,
[3]Department of Pharmacology, University of Patras Medical School

Lipidomics

Lipidomics is an emerging field of study of a cell's or organism's lipidome regarding its structure and function in addition to various interactions that may be present with other lipids or proteins. Under this definition, High-Density Lipoproteins, apolipoproteins and a variety of lipid classes were analyzed in order to extract underlying associations about their structural and function properties.

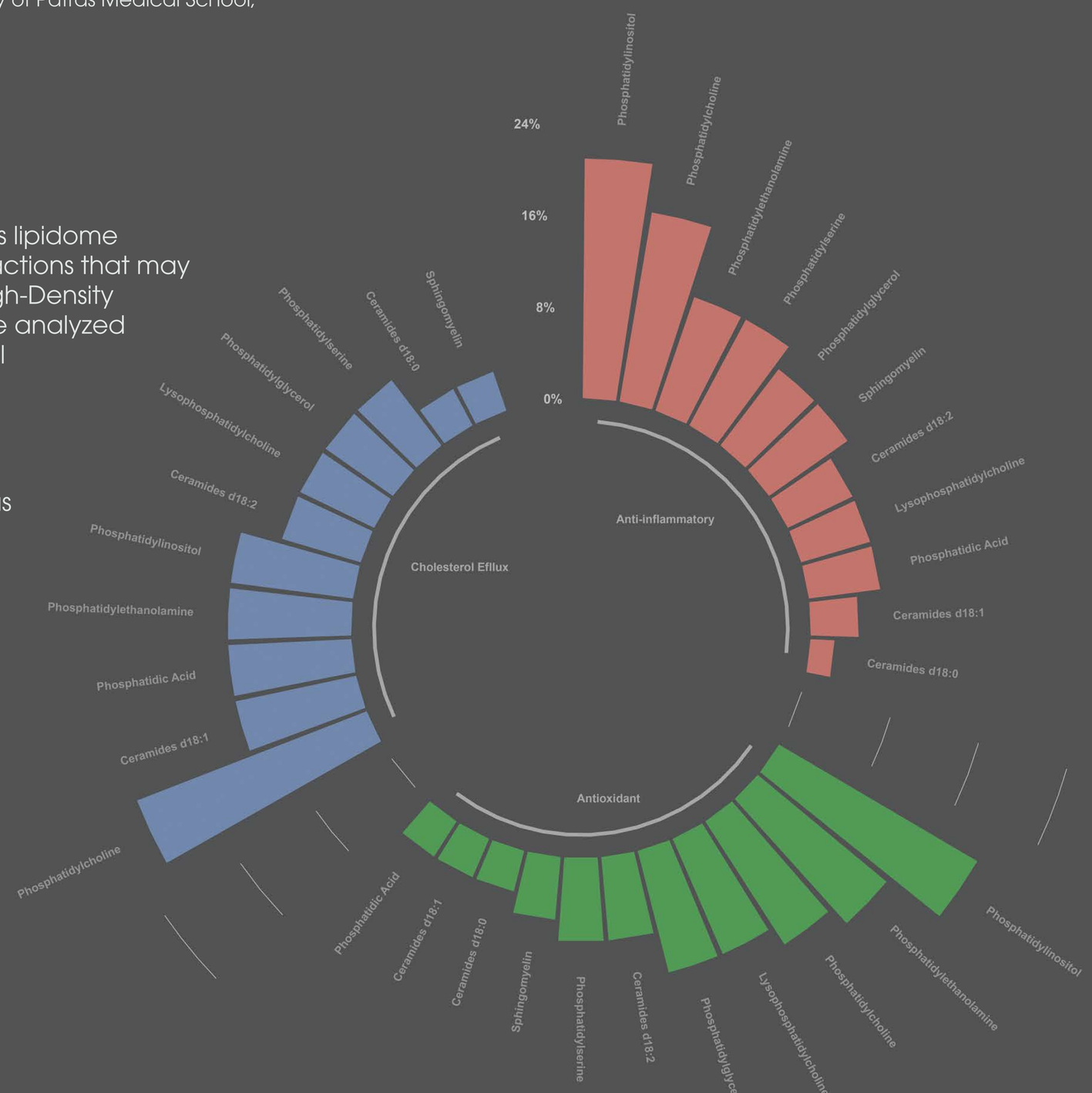
Research Objective

A framework for feature extraction with the use of lipids such as phospholipids, fatty acids, triglycerides and cholesterol and apolipoproteins such as APOA1, APOE3, and APOC3 in an effort to understand the underlying association between them and three major biological functions of HDL:

Reverse Cholesterol Transport
Anti-inflammatory Function
Antioxidant Function

Data Source

Part of data and other information have originated from the combination of two research publications, (1)(2) both regarding the analysis of HDL and some of its subpopulations in specific genetic background mice (C57BL/6 and Apoe-/- x Apoa1-/-, respectively) infected with adenoviruses expressing three different human apolipoproteins, APOC3 and APOA1, APOE3, respectively



Percentage Effect of Lipid Subclasses by HDL Function

Framework

Data Cleaning
Removal of variables with low variance, data transformations, etc.

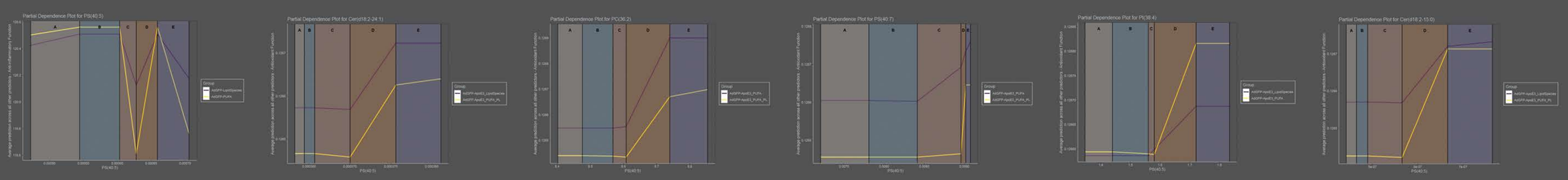
Exploratory Analysis
Hypothesis group testing, descriptive statistics on both datasets

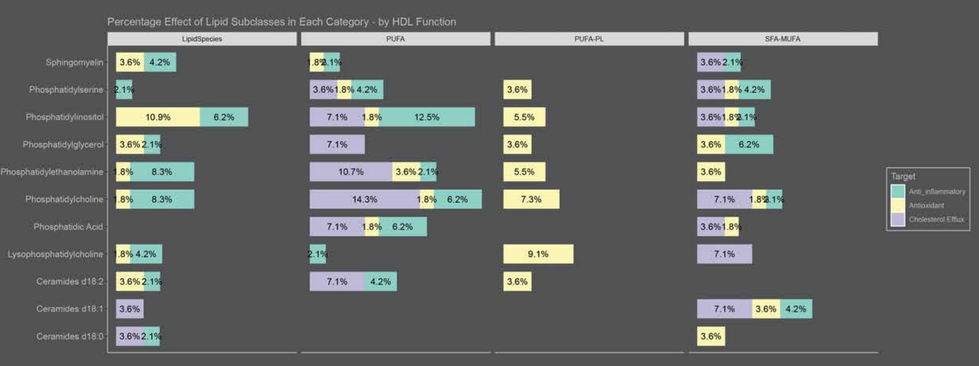
Random Forests
3 target models / category of interest x 2 datasets =24 trained models
Filtered variable importance plots based on Percent Increase in Mean Square Error (%IncMSE)
Values indicating this threshold were calculated by taking the absolute value of the highest negative ranked variable in %IncMSE and then select only positive ranked variables that pass the respective threshold.

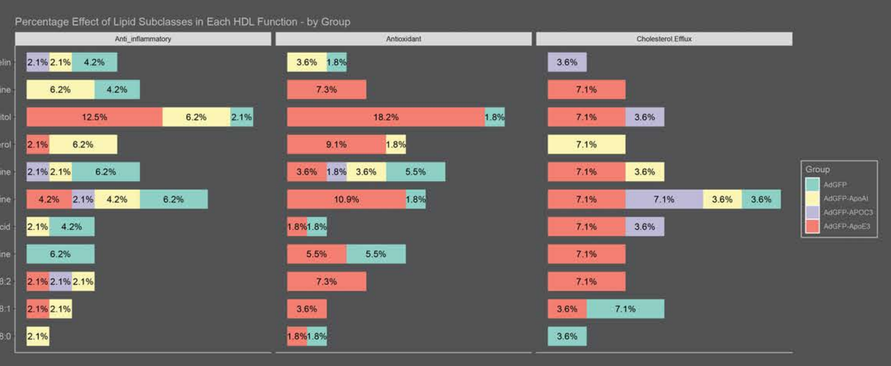
Partial Dependence Plots
An "under the hood" view of each model to determine effect ranges on each HDL function

Results

Phosphatidylinositols were presented as a major factor with high predictive power regarding both Anti-inflammatory and Antioxidant functions (20.8% and 20% total important fatty acids respectively). Phosphatidylcholines have the highest participation when it comes in predicting Cholesterol Efflux with 21.4%, 18.7% in Anti-inflammatory function and a 12.7% of total lipid subclasses in Antioxidant function.
Exploring common fatty acids individually,revealed that the majority derives from the AdGFP-ApoE3 group (RMSE = 0.0062, R^2 = 0.45), when the antioxidant function (0.13±0.01 (mean±SD)) is used as a target variable, showcasing the dominant effect of apolipoprotein E3 over other apolipoproteins.







Conclusion

A robust and scalable framework for feature extraction regarding lipidomics data is proposed based on benchmarks conducted on both p>>n and big data.

References

(1) Zvintzou, E., et al., Pleiotropic effects of apolipoprotein C3 on HDL functionality and adipose tissue metabolic activity. J Lipid Res, 2017.58(9): p. 1869-1883.
(2) Filou, S., et al., Distinct Roles of Apolipoproteins A1 and E in the Modulation of High-Density Lipoprotein Composition and Function. Biochemistry, 2016. 55(27): p. 3752-62.
(3) B. M. Greenwell, "pdp: An R Package for Constructing Partial Dependence Plots," The R Journal, vol. 9, no. 1, pp. 421-436, 2017.
(4) T. Hastie, R. Tibshirani and J. Friedman, The Elements of Statistical Learning, New York, NY: Springer New York, 2009