**Supplementary Material for Grouping of Complex Substances Using Analytical Chemistry Data: A Framework for Quantitative Evaluation and Visualization**

Melis Onel1,2, Burcu Beykal1,2, Kyle Ferguson3, Weihsueh A. Chiu3, Thomas J. McDonald4,

Lan Zhou5, John S. House6, Fred A. Wright6,7, David A. Sheen8,

Ivan Rusyn3,&\*, Efstratios N. Pistikopoulos1,2,&\*

1Artie McFerrin Department of Chemical Engineering, Texas A&M University, College Station, TX 77843, United States of America

2Texas A&M Energy Institute, Texas A&M University, College Station, TX, United States of America

3Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, TX, United States of America

4Department of Environmental and Occupational Health, Texas A&M University, College Station, TX, United States of America

5Department of Statistics, Texas A&M University, College Station, TX, United States of America

6Bioinformatics Research Center, North Carolina State University, Raleigh, NC, United States of America

7Departments of Statistics and Biological Sciences, North Carolina State University, Raleigh, NC

8Chemical Sciences Division, National Institute of Standards and Technology, Gaithersburg, MD, United States of America

& These authors contributing equally to this work.

\*corresponding authors

irusyn@cvm.tamu.edu

stratos@tamu.edu

The R Markdown documentation of unsupervised and supervised analysis workflow can be accessed through <http://parametric.tamu.edu/research/Onel_etAl_2019_Rmarkdown.html>

The GC-MS, GCxGC-FID, and IM-MS data of Petroleum UVCB, and GC-MS data of SRM samples can be downloaded from http://paroc.tamu.edu/Software/Onel\_etAl\_2019\_data.zip