Supporting information

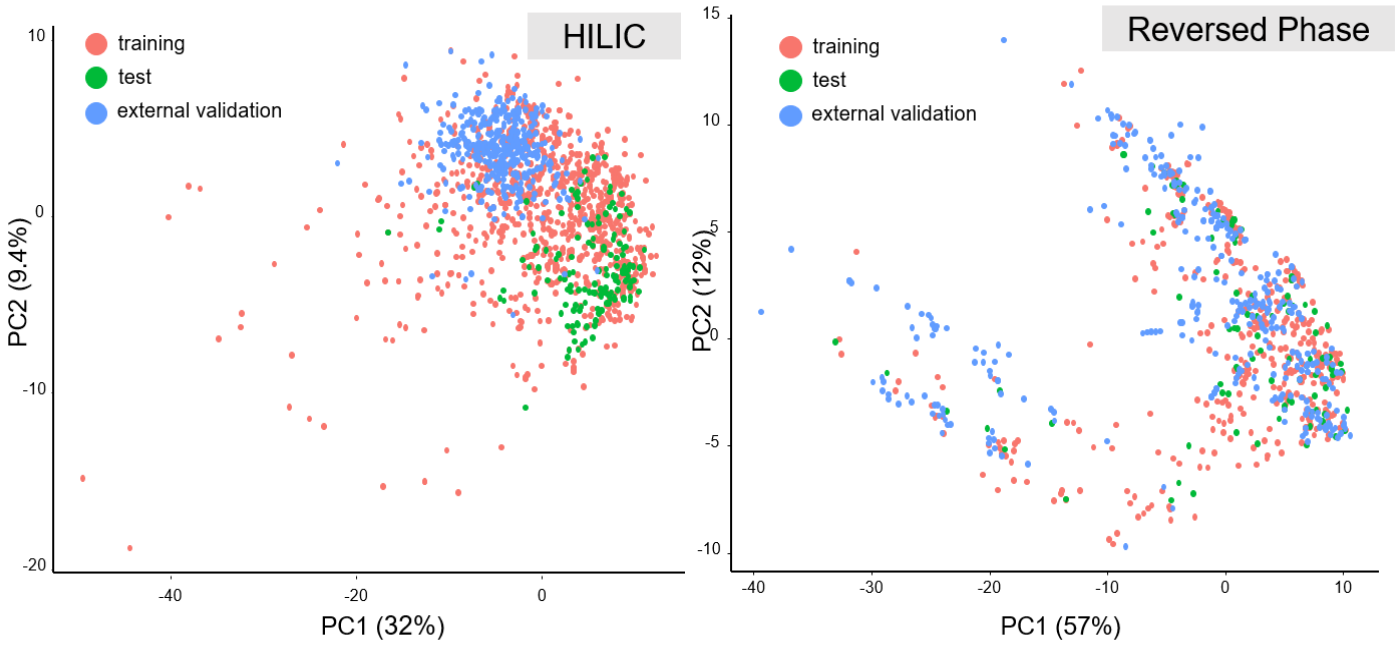
Retip: retention time prediction for compound annotation in untargeted metabolomics

Paolo Bonini[(1)*], Tobias Kind[(2)], Hiroshi Tsugawa[(3,4)] Dinesh Kumar Barupal[(2)] and Oliver Fiehn[(2)*]

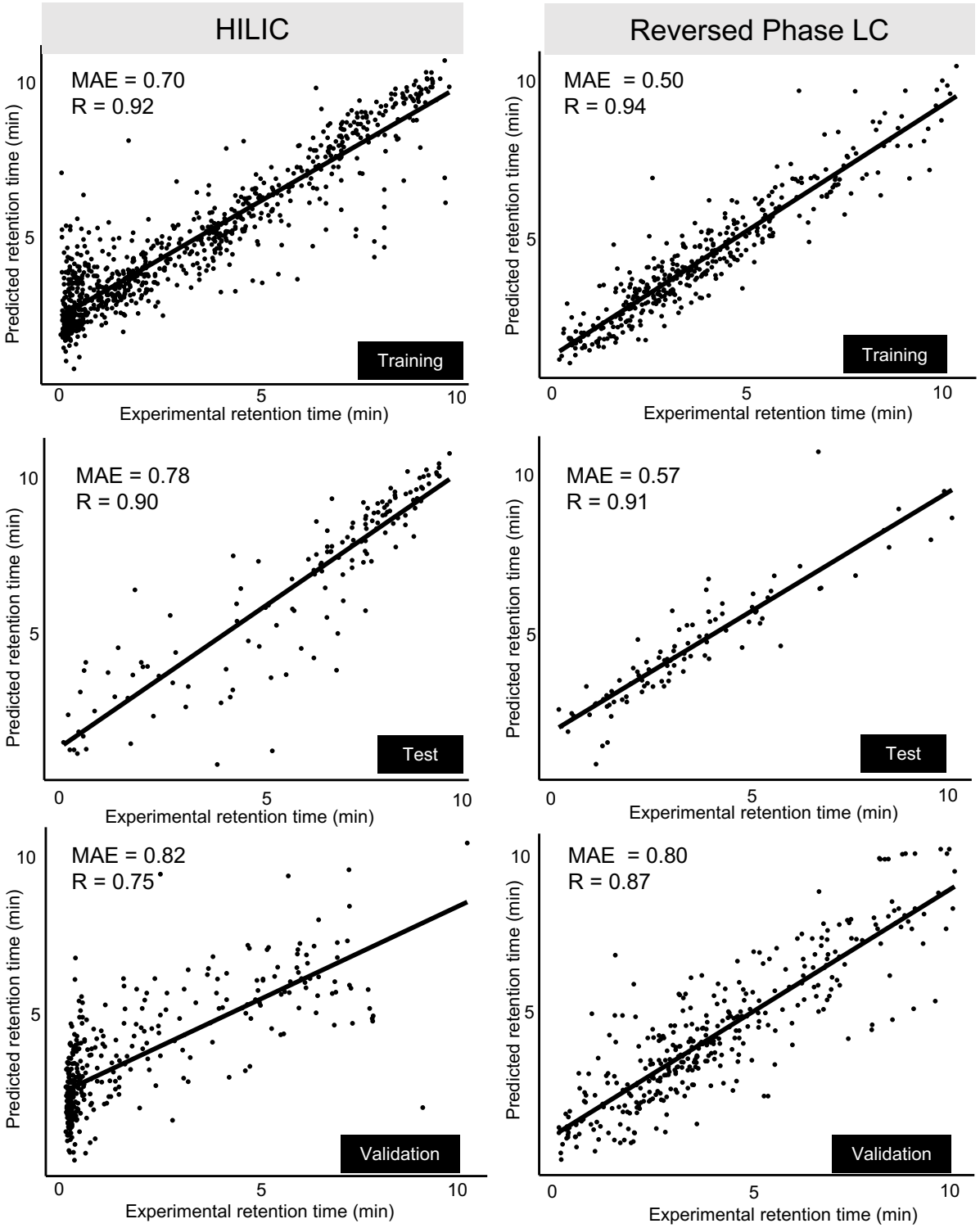
(1) NGAlab, La Riera de Gaia, Tarragona 43762, Spain (2) West Coast Metabolomics Center, UC Davis Genome Center, University of California, Davis, 451 Health Sciences Drive, Davis, California 95616, United States (3) RIKEN Center for Sustainable Resource Science, Yokohama, Japan (4) RIKEN Center for Integrative Medical Sciences, Yokohama, Japan

corresponding author email ofiehn@ucdavis.edu

Table of contents



Supporting Figure S1. Predicting LC-retention times on HILIC- and Reversed-phase LC methods. Principal Component Analyses (PCA) on 2D-chemical descriptors to show chemical diversity in training and validation sets. 147 descriptors for HILIC and 142 descriptors for RP were used. Data were centered and auto-scaled.



Supporting Figure S2. Visualizing prediction errors for Keras machine learning for HILIC- and reversed-phase LC methods. MAE is mean absolute error, R is correlation coefficient for experimental versus predicted retention times. Keras method details are given in the main manuscript. Data from Supporting Tables S1-S6.