Additional File 1

Comprehensive Benchmarking of Computational Tools for Predicting Toxicokinetic and Physicochemical Properties of Chemicals

Domenico Gadaleta[1,*], Eva Serrano Candelas[2,*], Rita Ortega Vallbona[2] , Erika Colombo[1], Marina Garcia de Lomana[3], Giada Biava[1], Pablo Aparicio-Sánchez[2,†], Alessandra Roncaglioni[1], Rafael Gozalbes[2, 4], Emilio Benfenati[1]

AFFILIATIONS:

[1] Laboratory of Environmental Chemistry and Toxicology, Department of Environmental Health Sciences, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Milan, Italy

[2] ProtoQSAR SL, CEEI (Centro Europeo de Empresas Innovadoras), Valencia, Spain

[3] Bayer AG, Machine Learning Research, Research & Development, Pharmaceuticals, Germany

[4] Moldrug AI Systems SL, c/Olimpia Arozena Torres 45, Valencia 46018, Spain.

[*]Equally contributing first authors

[†] Current address: Spanish National Cancer Research Center (CNIO), Experimental Therapeutics Programme, Madrid, Spain

1. **Intra-dataset Outlier Detection Using Z-scores (Continuous Data): The figures present the intra-outlier analysis for the validation datasets associated to a specific property based on Z-scores:**
   - **Top row:** Z-scores for each dataset are plotted against the chemical ID before outlier removal. The red dashed lines indicate the Z-score threshold of ±3, with values outside this range considered as outliers ('intra-outliers').
   - **Second row:** Categorization of the dataset records before outlier removal. Values of 1 represent identified intra-outliers, 0 represents non-outliers, and -1 corresponds to missing data in the specific validation dataset.
   - **Second row:** Z-scores for each dataset are plotted against the chemical ID before outlier removal. The red dashed lines indicate the Z-score threshold of ±3, showing cleaned datasets.
   - **Fourth row:** Categorization of the dataset records after outlier removal. Values of 1 represent identified intra-outliers, 0 represents non-outliers, and -1 corresponds to missing data in the specific validation dataset.

2. **Inter-dataset Outlier Detection Based on Standardized Deviation (Continuous Data): The figures present the inter-outlier analysis, comparing each pair of validation datasets associated to a specific property and identifying value discrepancies based on standardized standard deviations (STD).**
   - **Top row:** Correlation plots for the experimental values shared between each pair of datasets before outlier removal. The red line represents the perfect correlation. The R2 value and the percentage of common molecules are indicated in each plot.
   - **Second row:** Standardized standard deviation (STD) plots for each dataset pair before outlier removal. The red dashed line marks the threshold of standardized STD = 0.2, with points above this line considered as outliers ('inter-outliers').
   - **Third row:** Correlation plots for the experimental values shared between each pair of datasets after outlier removal. The red line represents the perfect correlation. The R2 value and the percentage of common molecules are indicated in each plot.
   - **Fourth row:** Standardized standard deviation (STD) plots for each dataset pair after outlier removal. The red dashed line marks the threshold of standardized STD = 0.2, showing cleaned datasets.

3. **Inter-dataset Outlier Detection (Categorical Data): The figures present a categorical 'inter-outlier' analysis for the validation datasets associated to a specific property, identifying category inconsistencies between shared records.**
   - **Top panel:** Comparison of the experimental values for shared compounds across both datasets before outlier removal. A value of 1 indicates agreement where both datasets show positive results for the compound, while 0 indicates agreement where both datasets show negative results. A value of -1 represents inconsistent results between the datasets, marking these compounds as inter-outliers. The percentage and total count of inter-outliers are noted in the plot.
   - **Bottom panel:** Comparison of the experimental values for shared compounds across both datasets after outlier removal. A value of 1 indicates agreement where both datasets show positive results for the compound, while 0 indicates agreement where both datasets show negative results. The -1 category (inconsistent records) is eliminated, indicating a clean dataset.

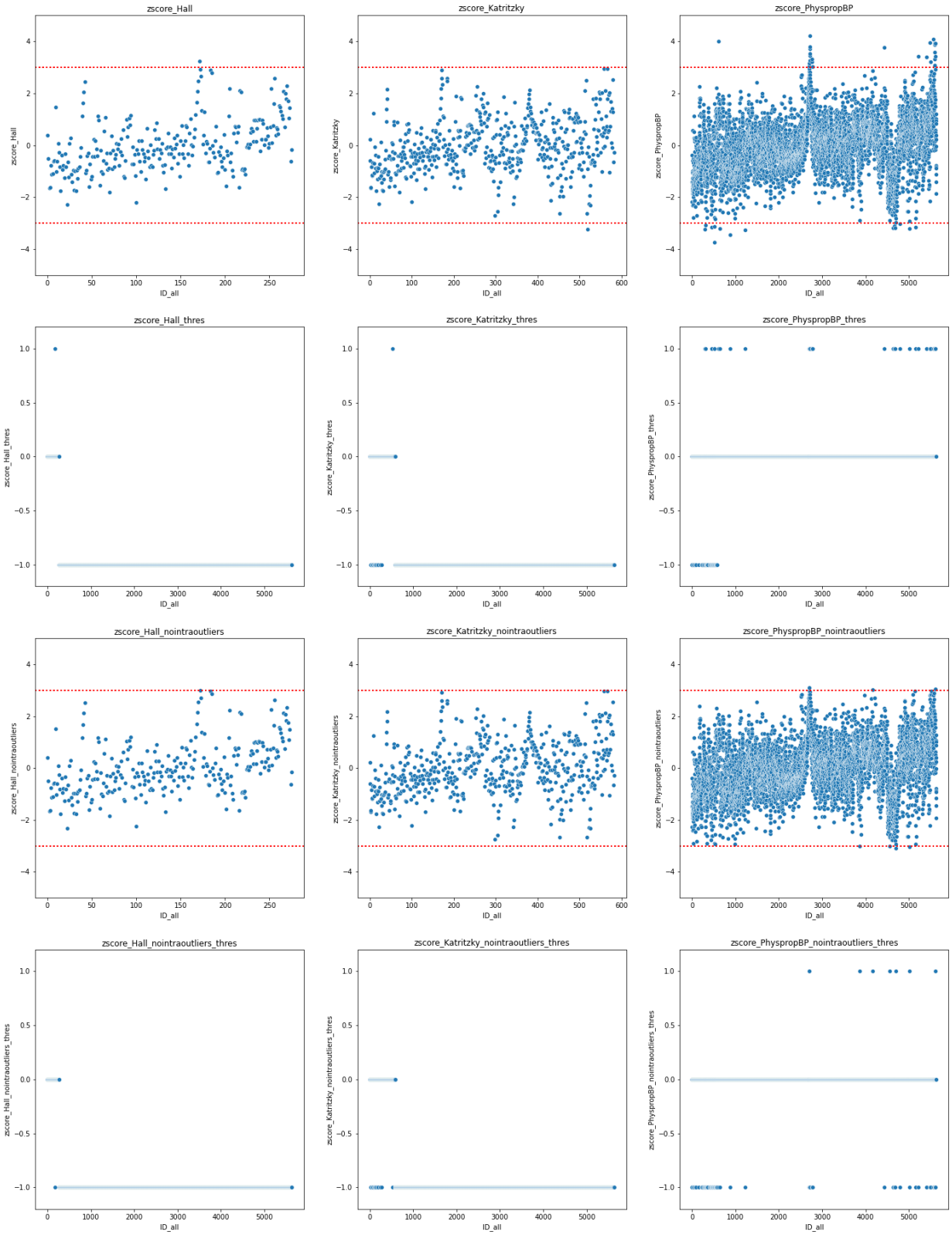


Figure S1: Boiling Point (BP). Intra-dataset Outlier Detection Using Z-scores (Continuous Data).

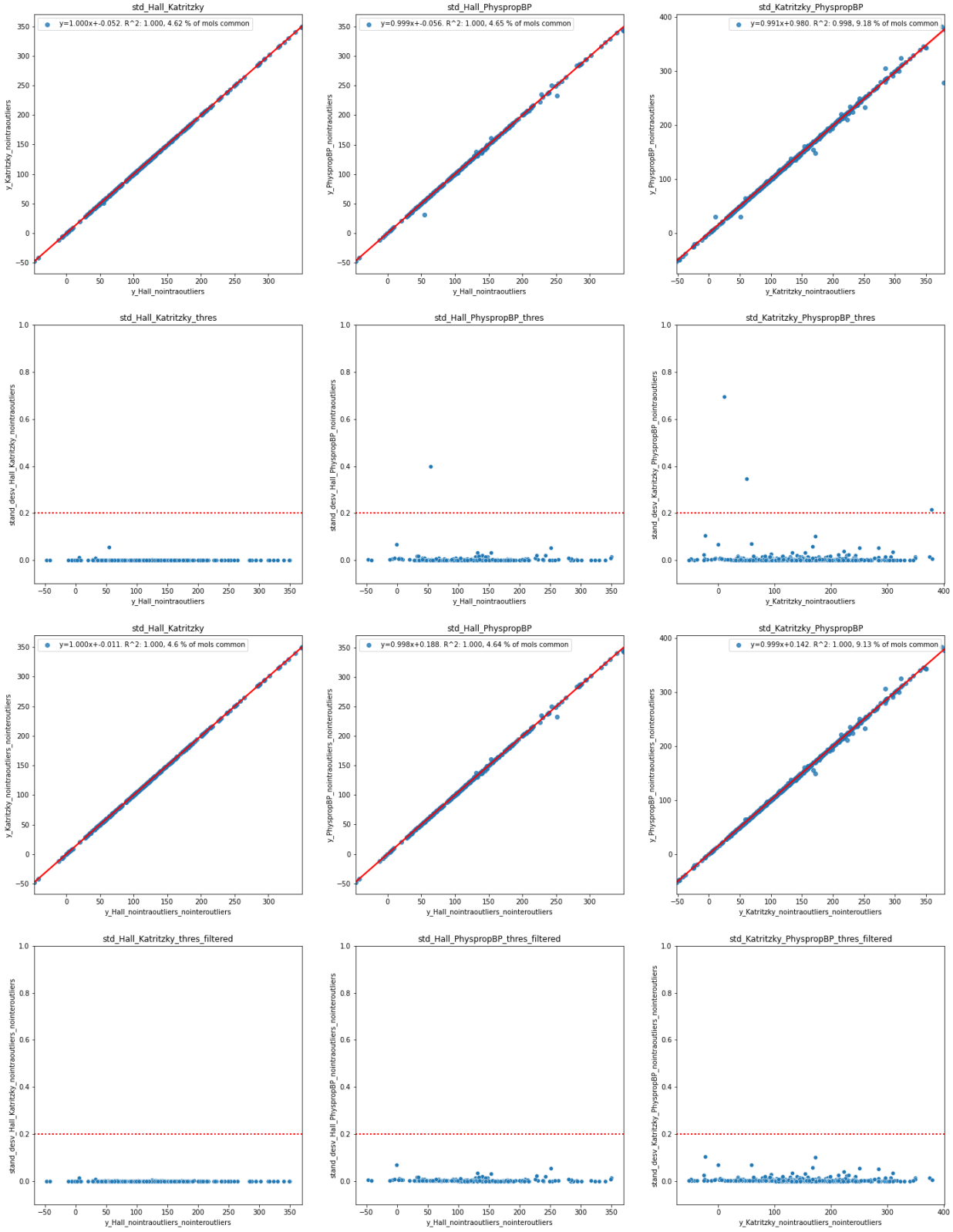


Figure S2: Boiling Point (BP). Inter-dataset Outlier Detection Based on Standardized Deviation (Continuous Data).

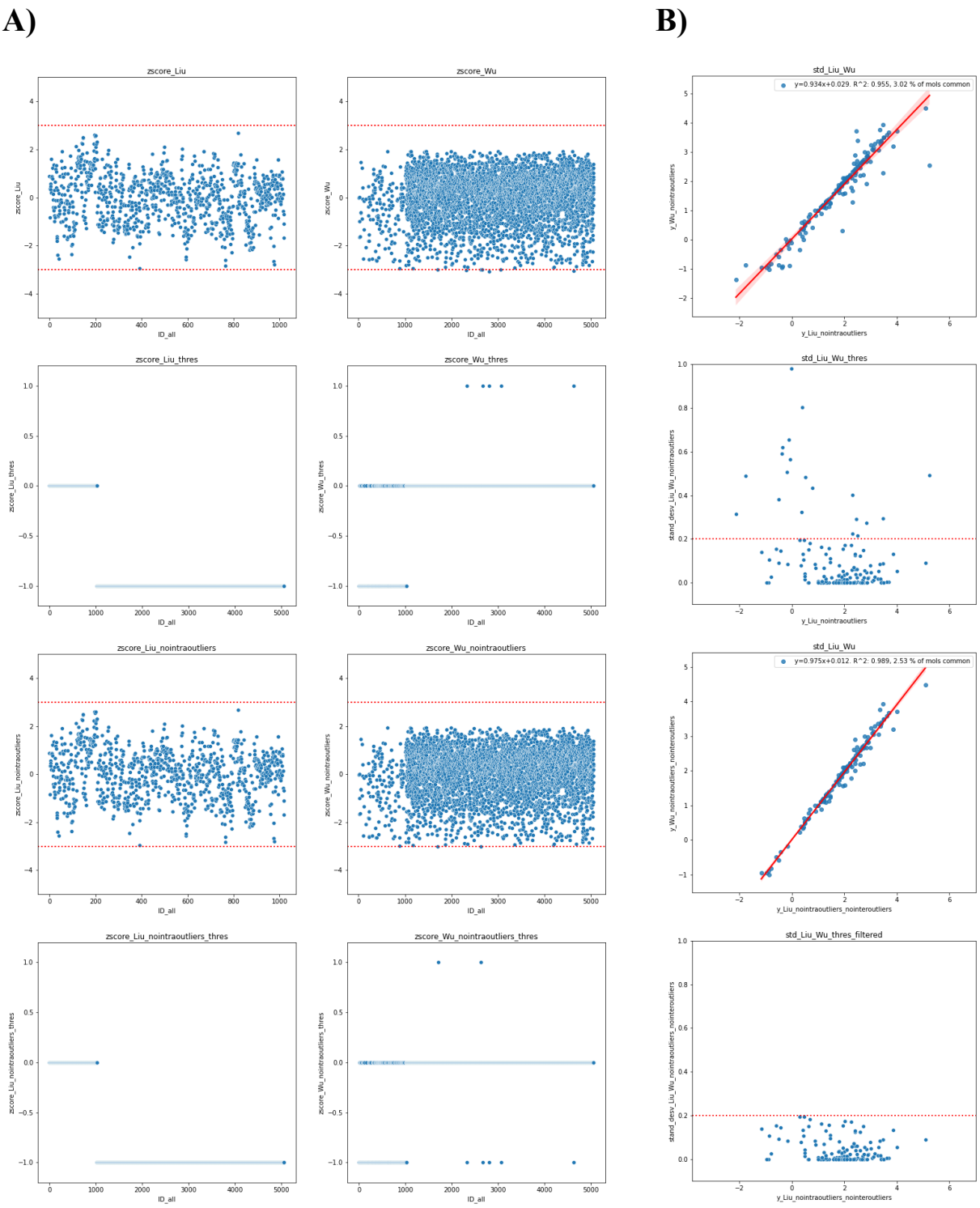


Figure S3: Distribution coefficient (logD). (A) Intra-dataset Outlier Detection Using Z-scores (Continuous Data); (B) Inter-dataset Outlier Detection Based on Standardized Deviation (Continuous Data).

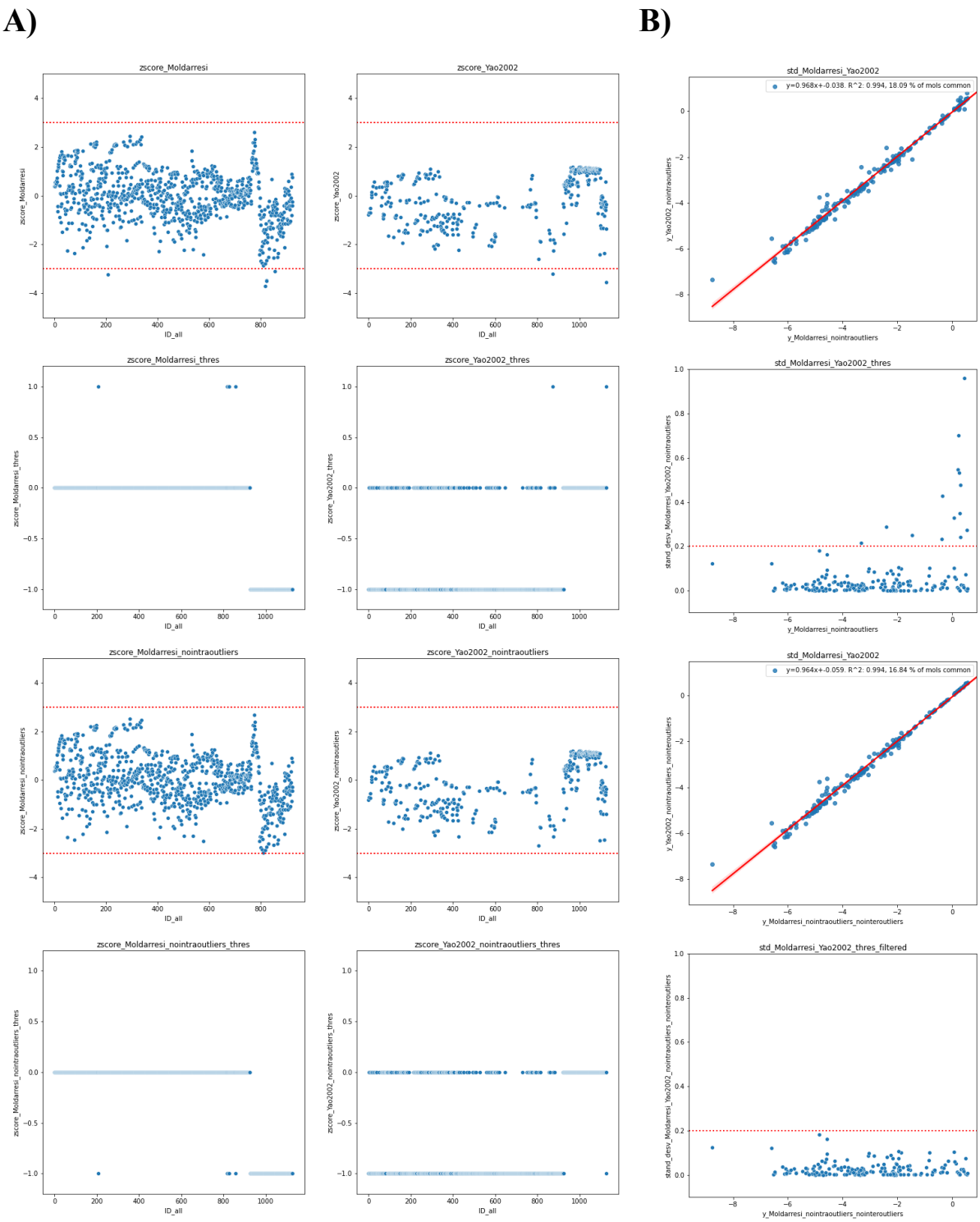


Figure S4: Henry's law constant (logH). (A) Intra-dataset Outlier Detection Using Z-scores (Continuous Data); (B) Inter-dataset Outlier Detection Based on Standardized Deviation (Continuous Data).

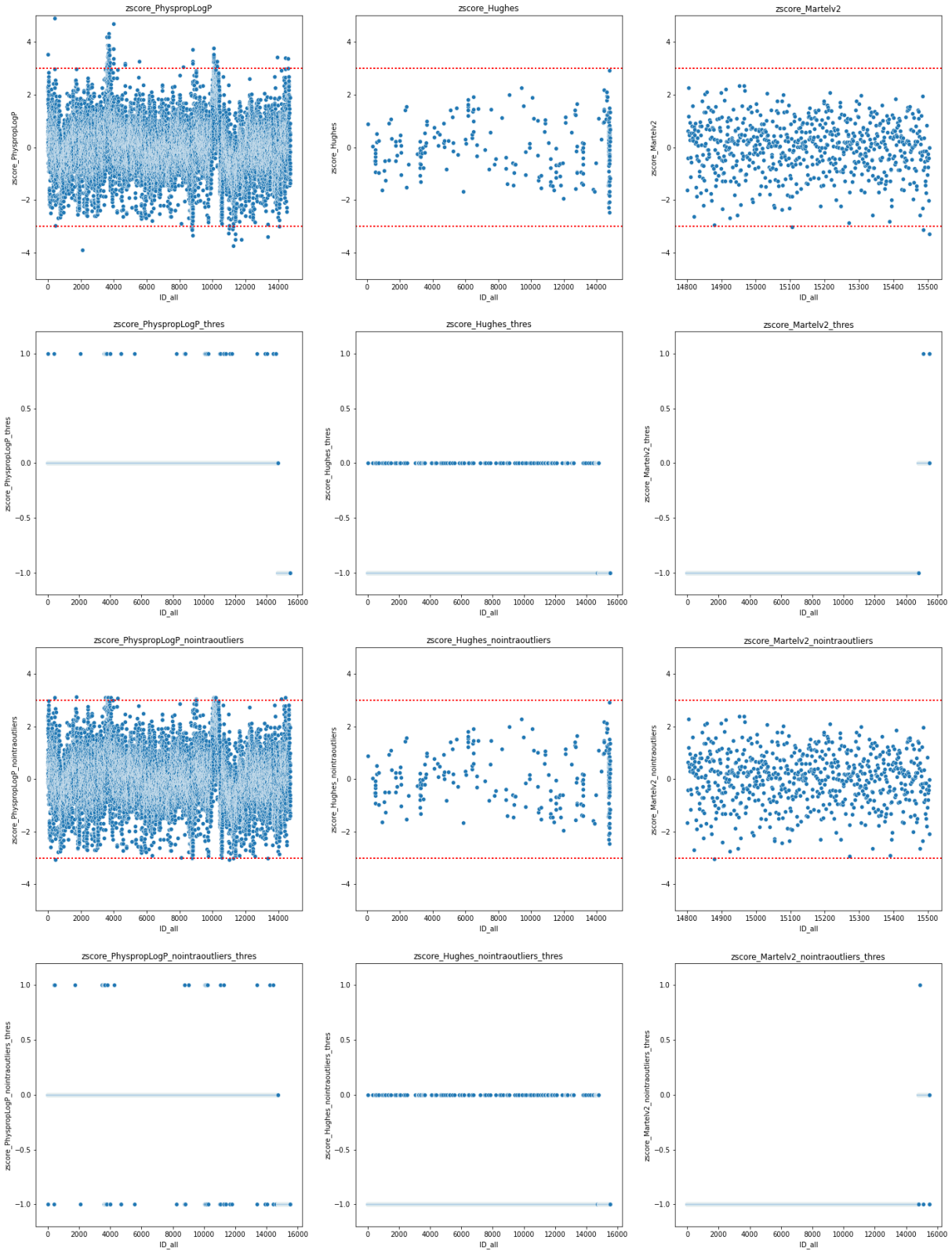


Figure S5: Octanol/water partition coefficient (logP). Intra-dataset Outlier Detection Using Z-scores (Continuous Data).

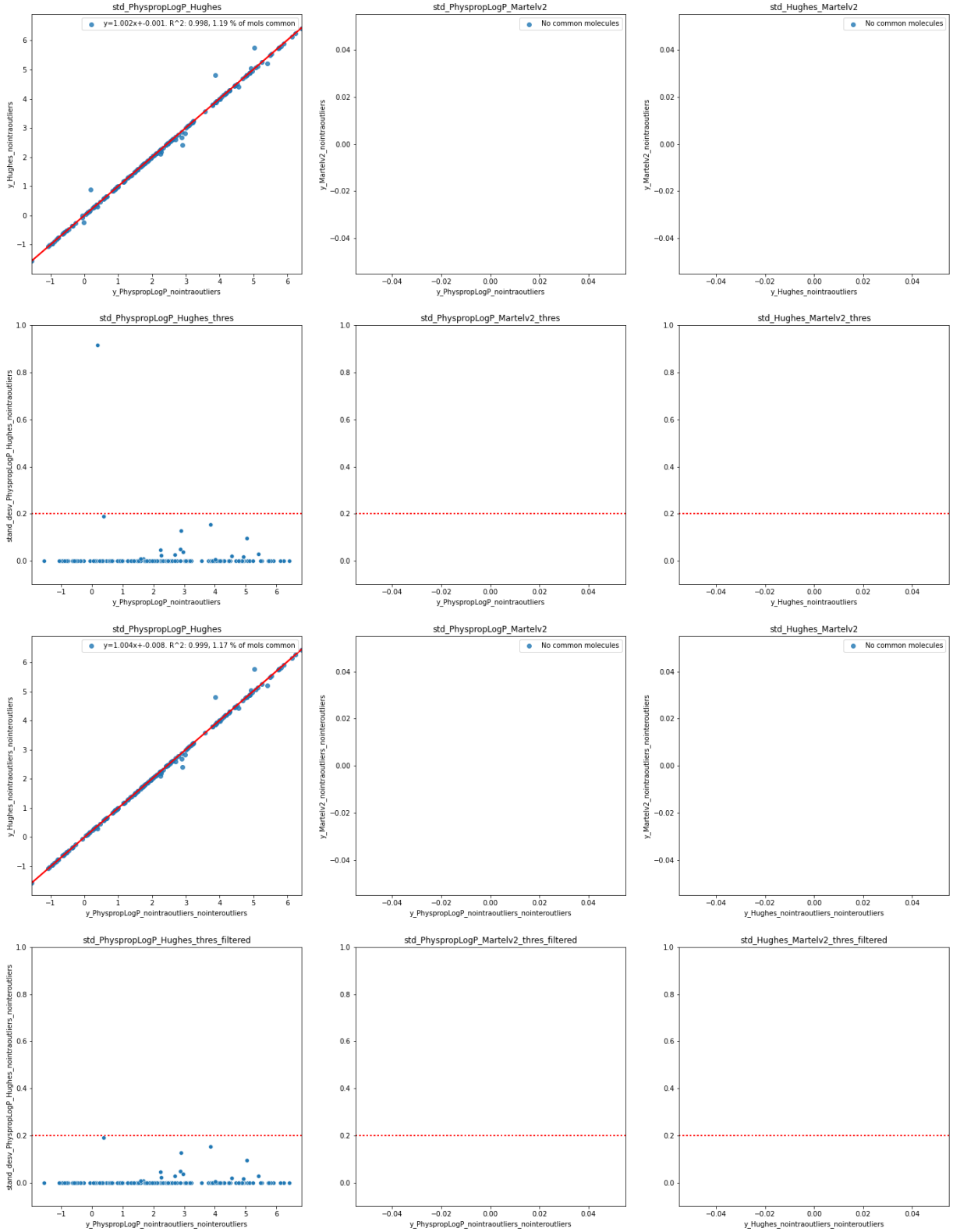
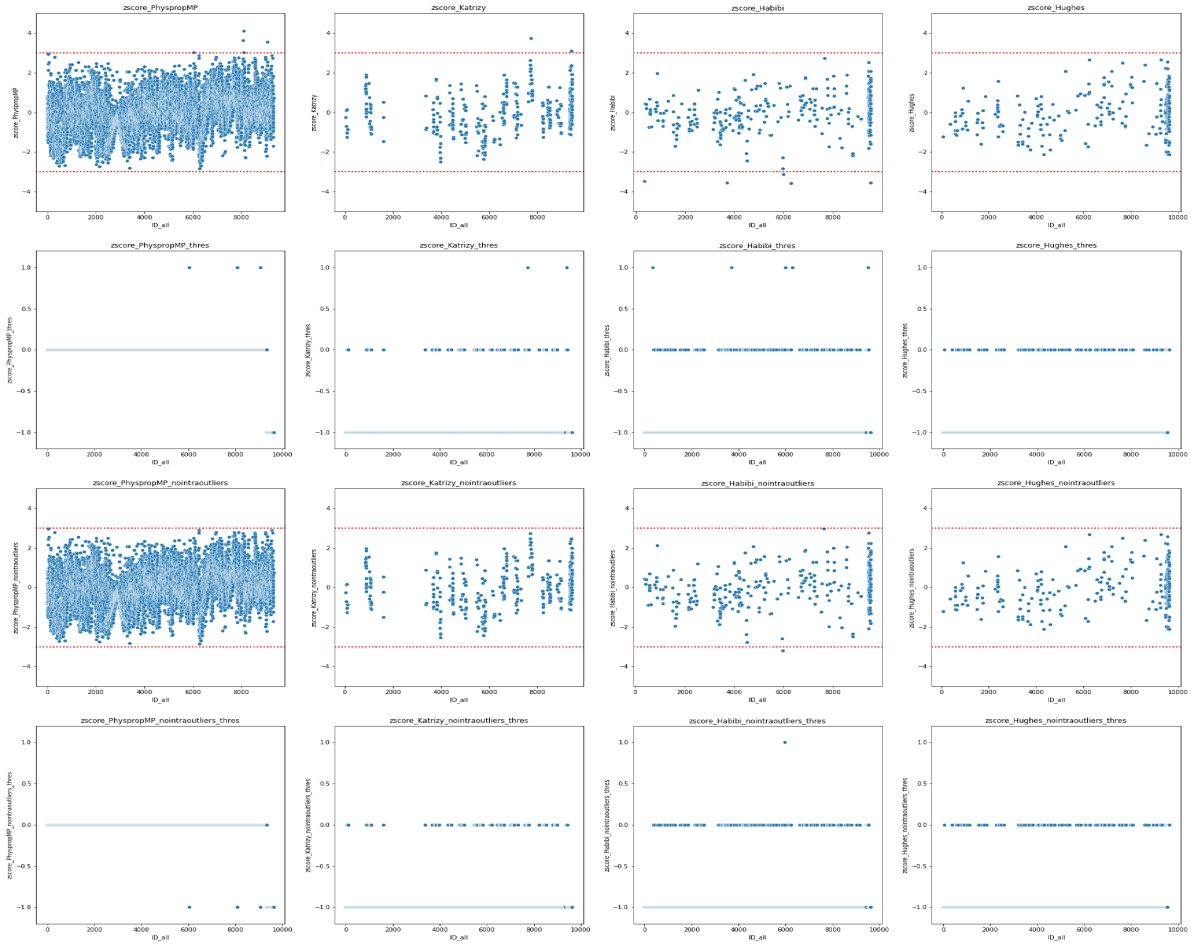


Figure S6: Octanol/water partition coefficient (logP). Inter-dataset Outlier Detection Based on Standardized Deviation (Continuous Data).

A)



B)

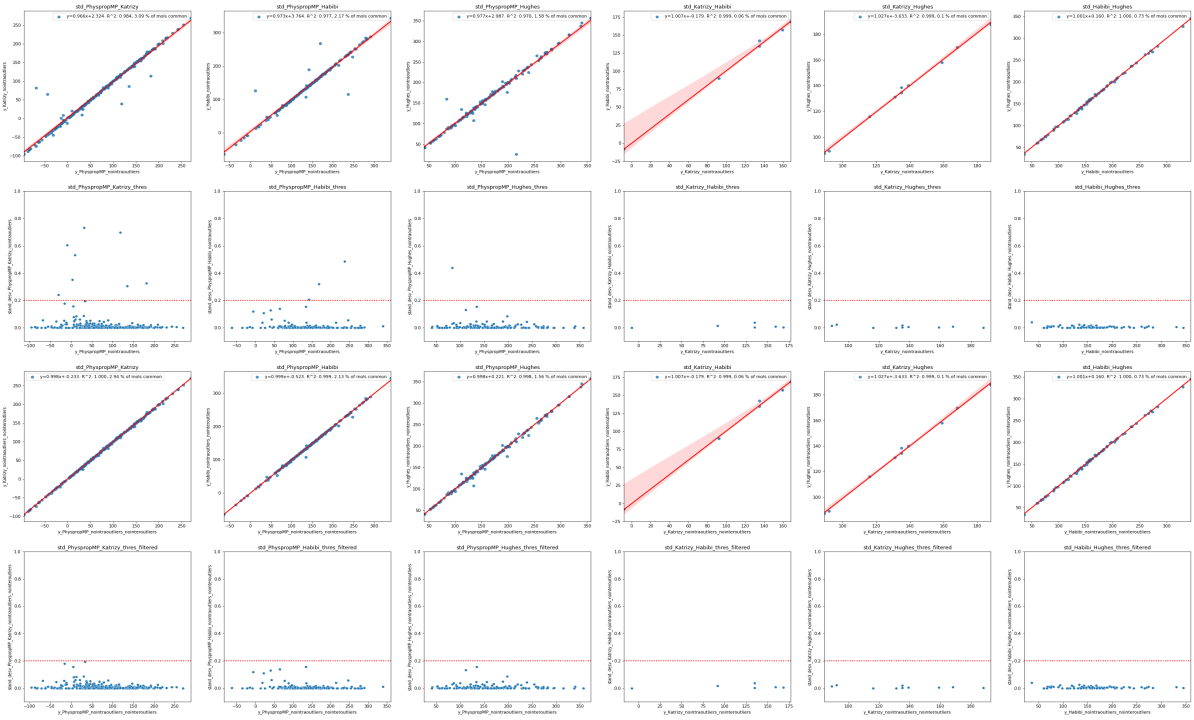
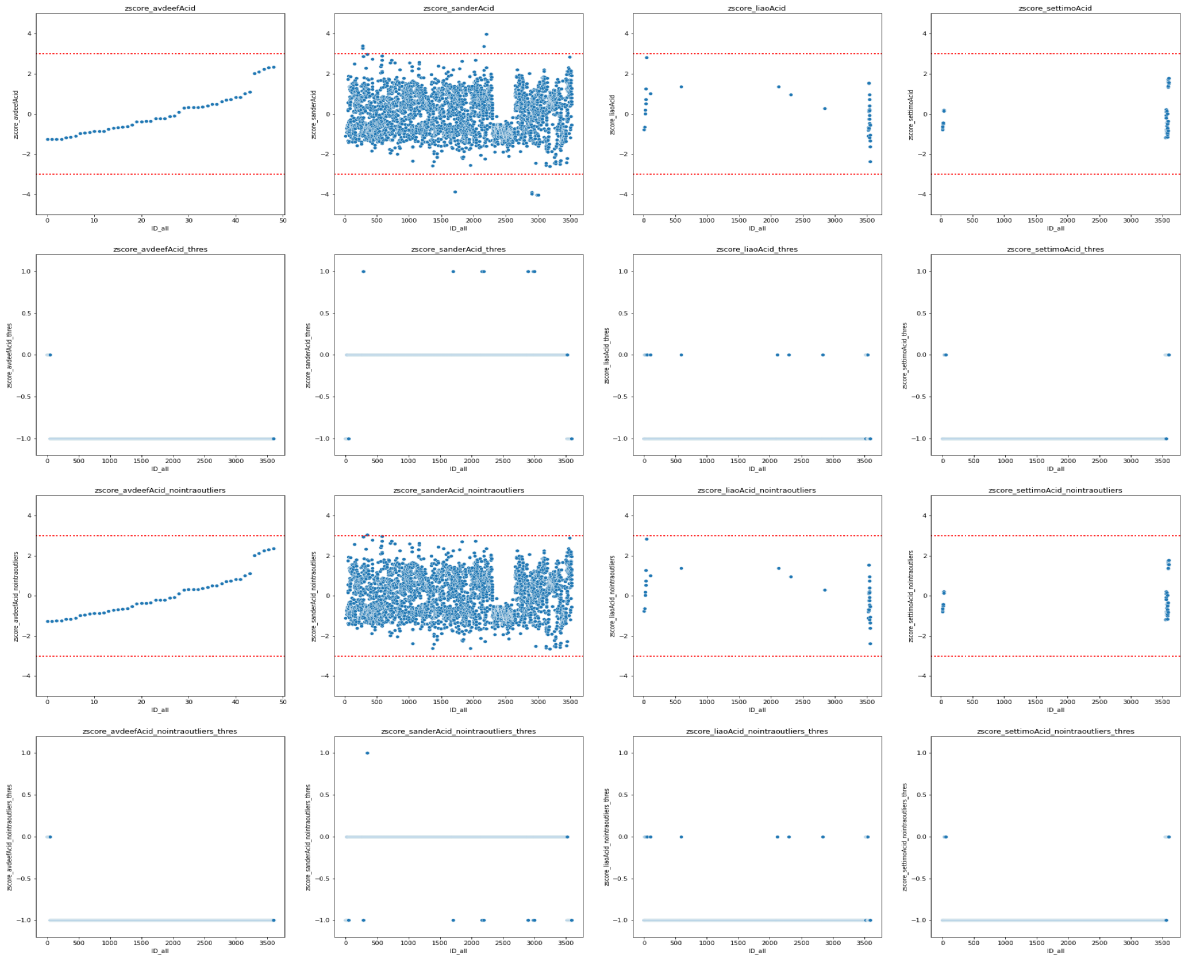


Figure S7: Melting Point (MP). (A) Intra-dataset Outlier Detection Using Z-scores (Continuous Data); (B) Inter-dataset Outlier Detection Based on Standardized Deviation (Continuous Data).

A)



B)

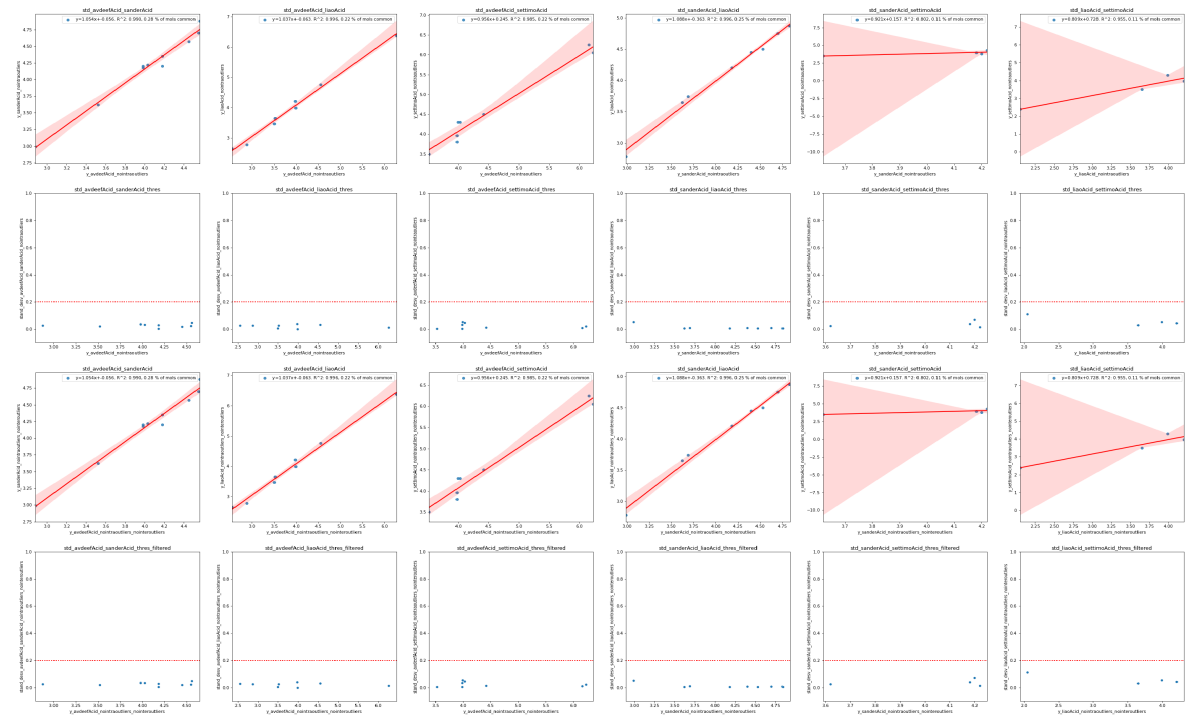
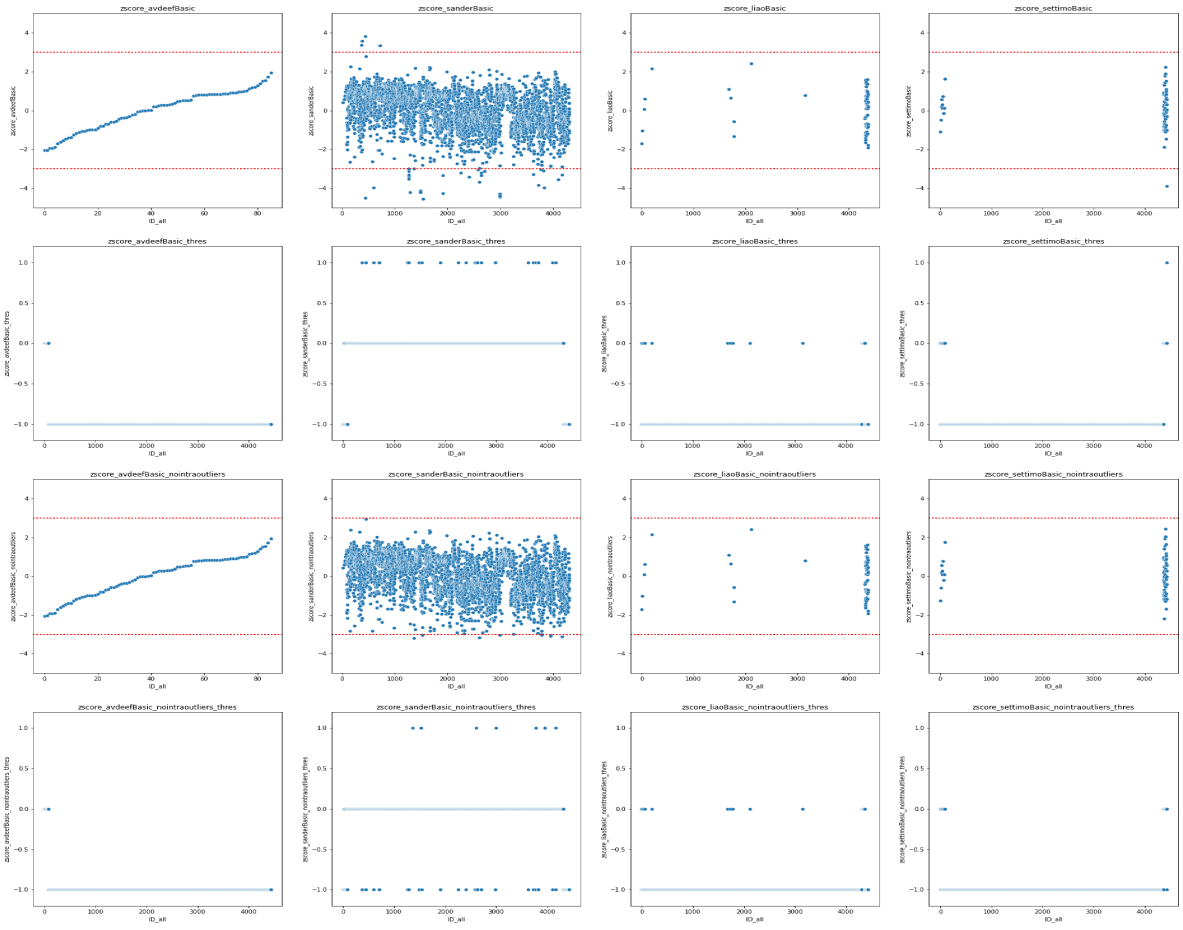


Figure S8: pKa acidic (pka-a). (A) Intra-dataset Outlier Detection Using Z-scores (Continuous Data); (B) Inter-dataset Outlier Detection Based on Standardized Deviation (Continuous Data).

A)



B)

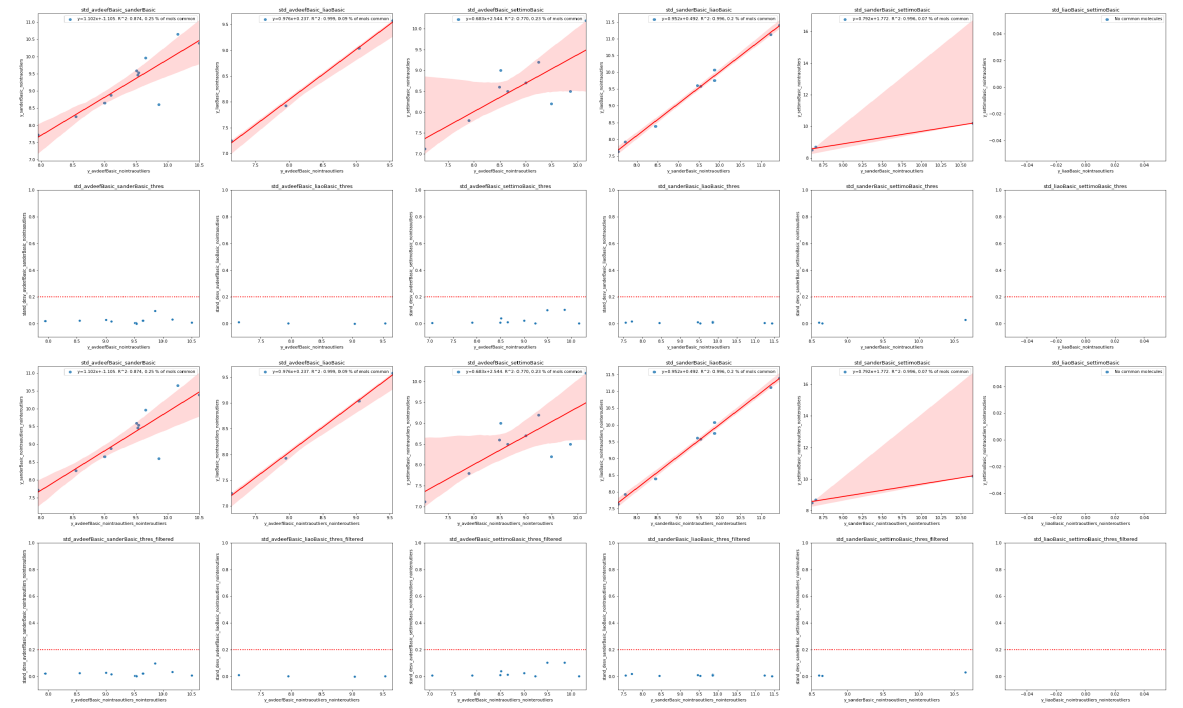


Figure S9: pKa basic (pka-b). (A) Intra-dataset Outlier Detection Using Z-scores (Continuous Data); (B) Inter-dataset Outlier Detection Based on Standardized Deviation (Continuous Data).

A) B)

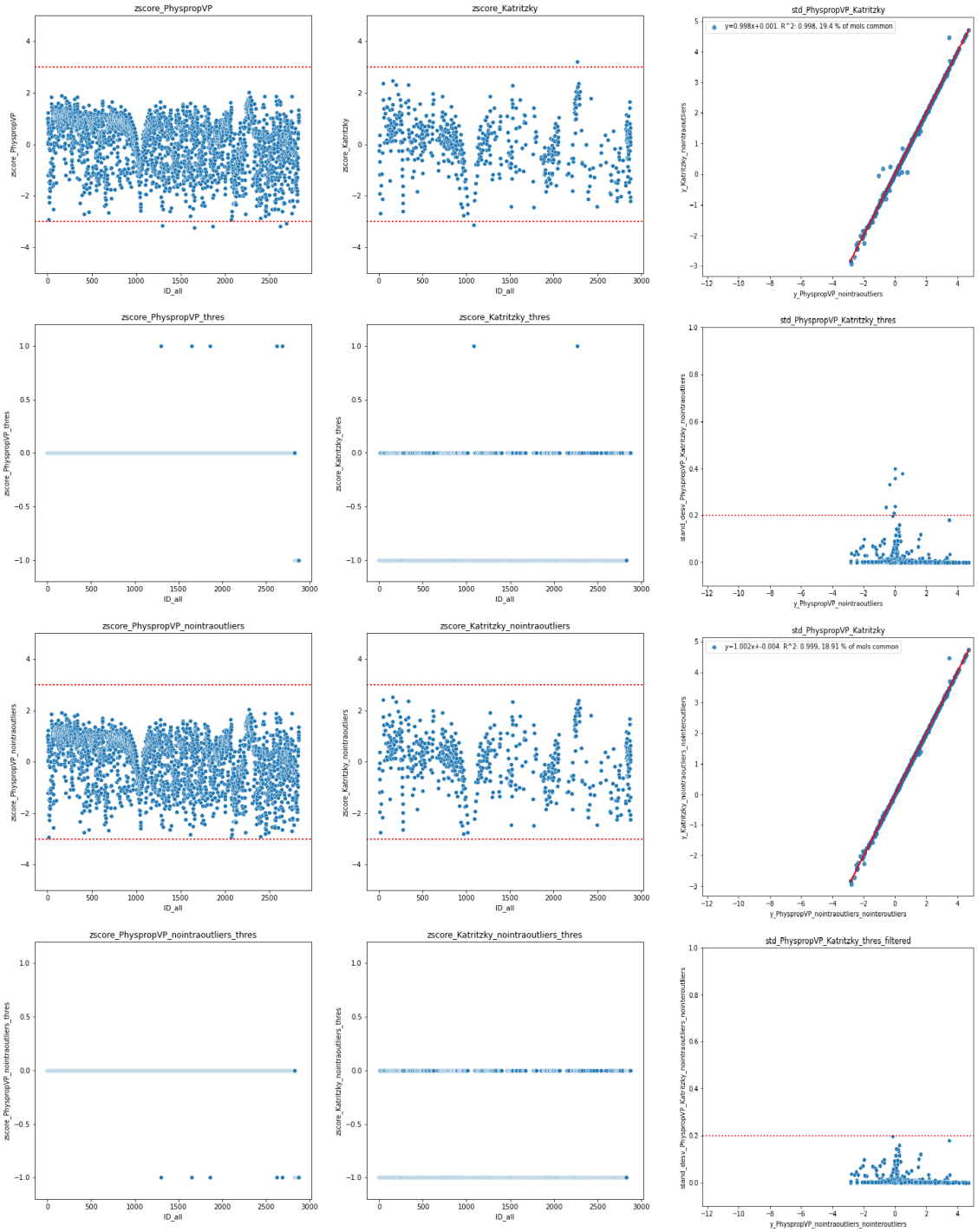


Figure S10: Vapor Pressure (logVP). (A) Intra-dataset Outlier Detection Using Z-scores (Continuous Data); (B) Inter-dataset Outlier Detection Based on Standardized Deviation (Continuous Data).

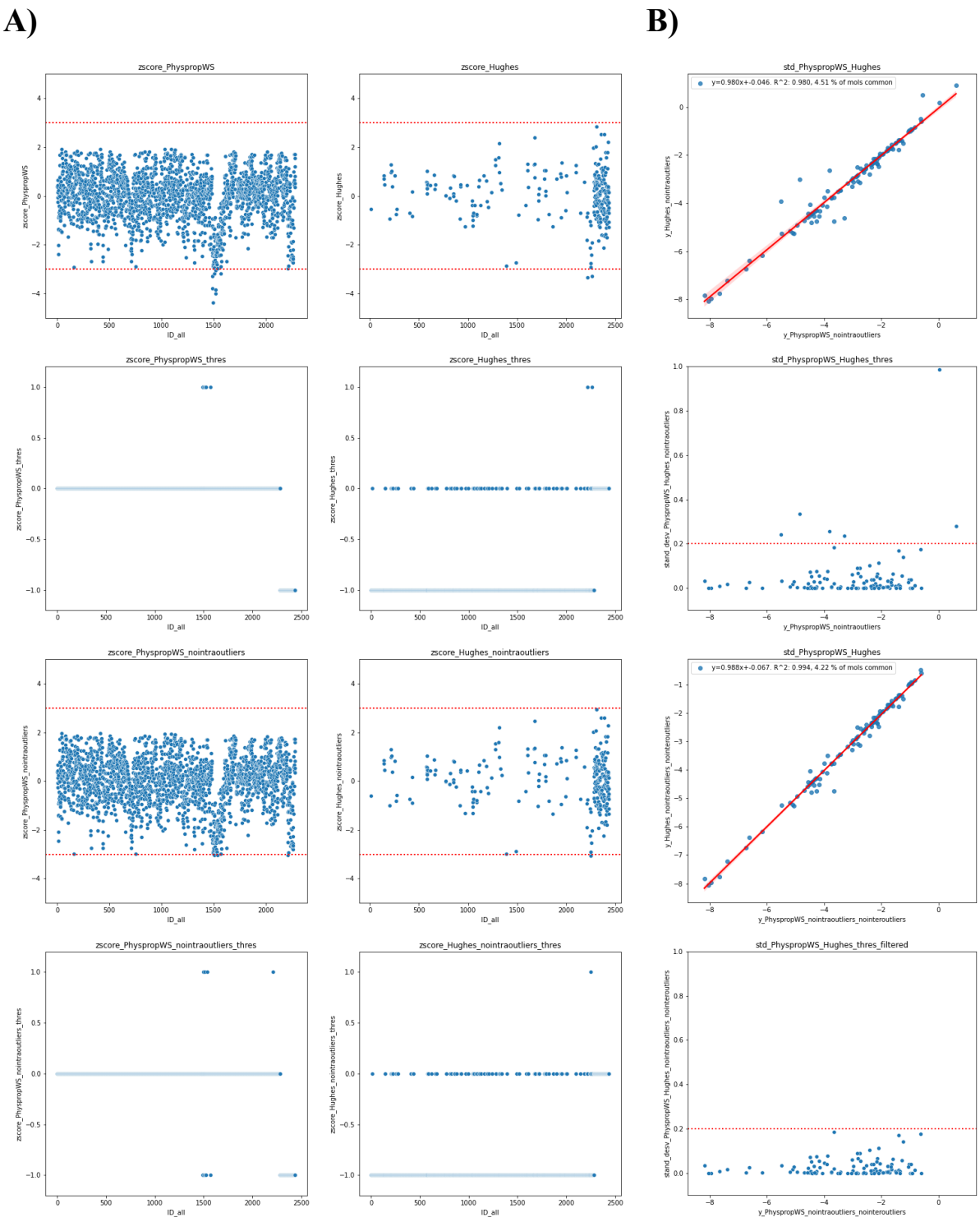


Figure S11: Water Solubility (logWS). (A) Intra-dataset Outlier Detection Using Z-scores (Continuous Data); (B) Inter-dataset Outlier Detection Based on Standardized Deviation (Continuous Data).

A) **B)**

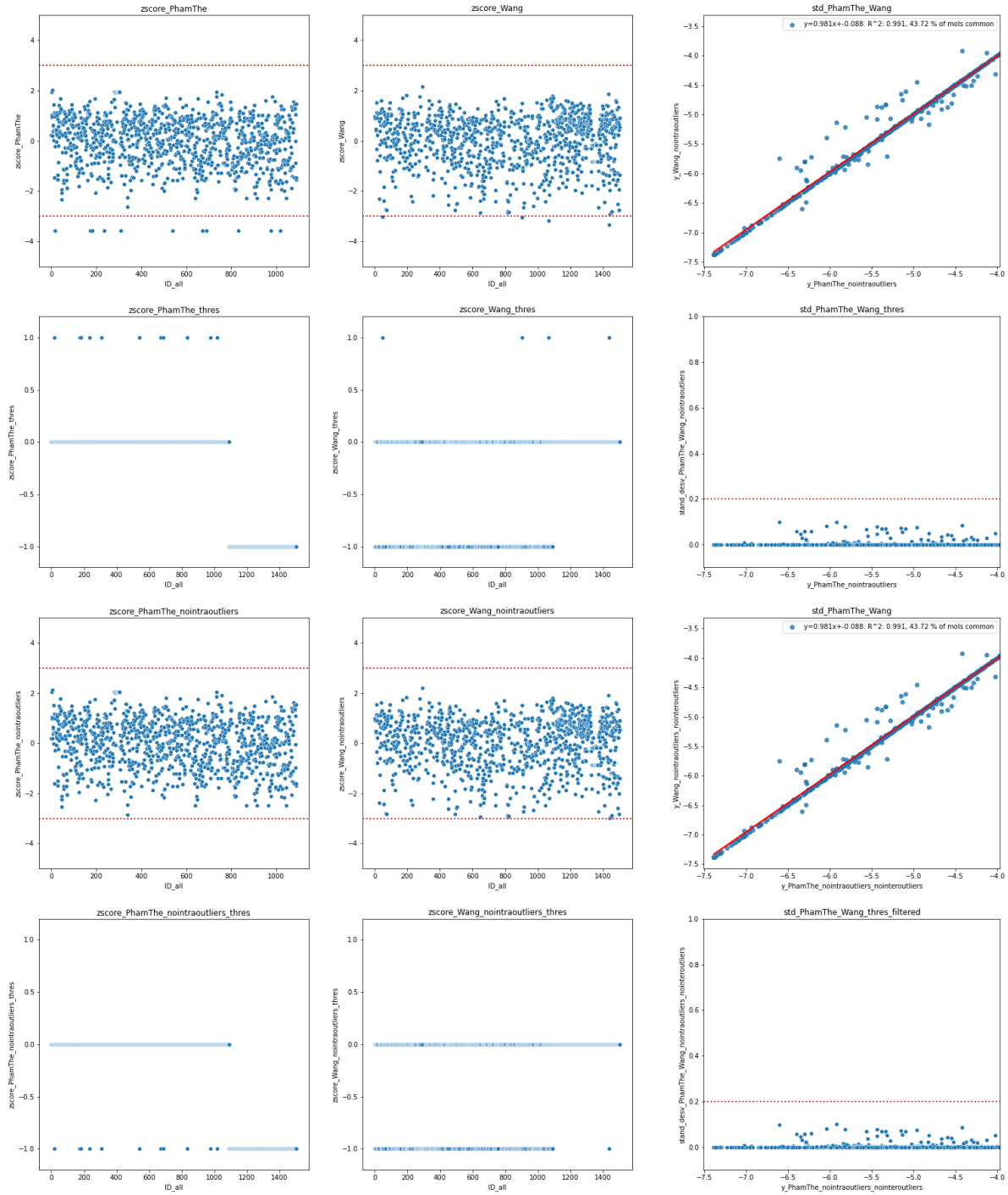


Figure S12: Caco-2 permeability (Caco2). (A) Intra-dataset Outlier Detection Using Z-scores (Continuous Data); (B) Inter-dataset Outlier Detection Based on Standardized Deviation (Continuous Data).

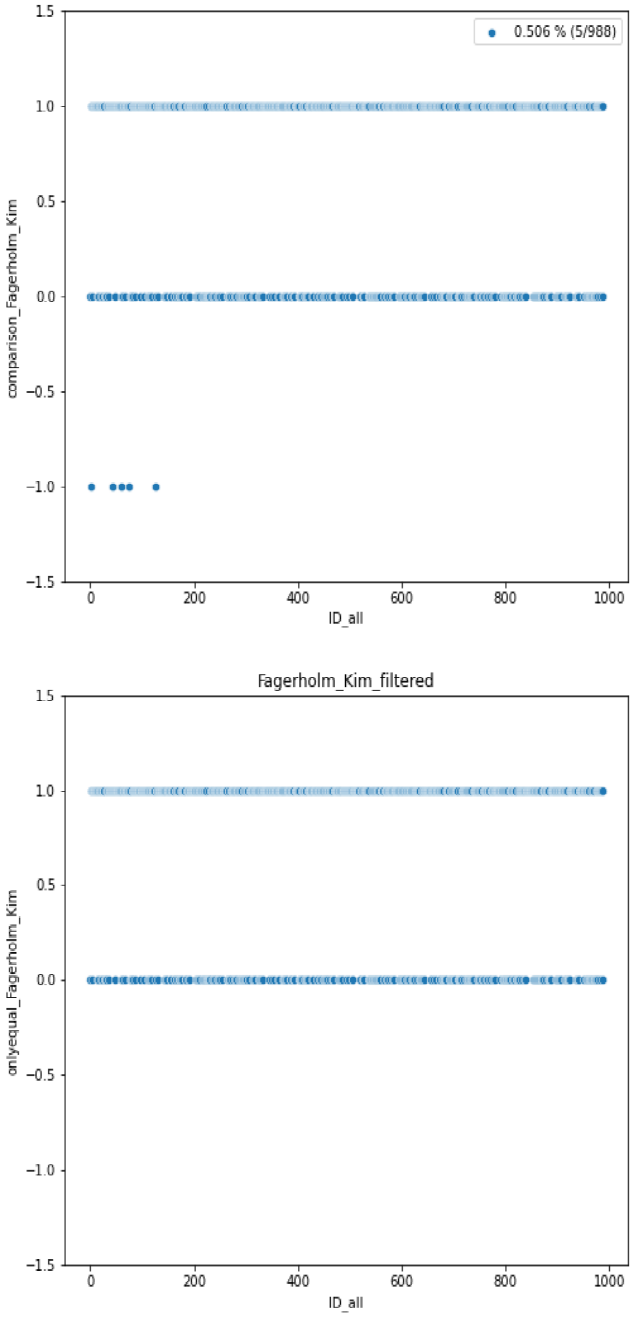
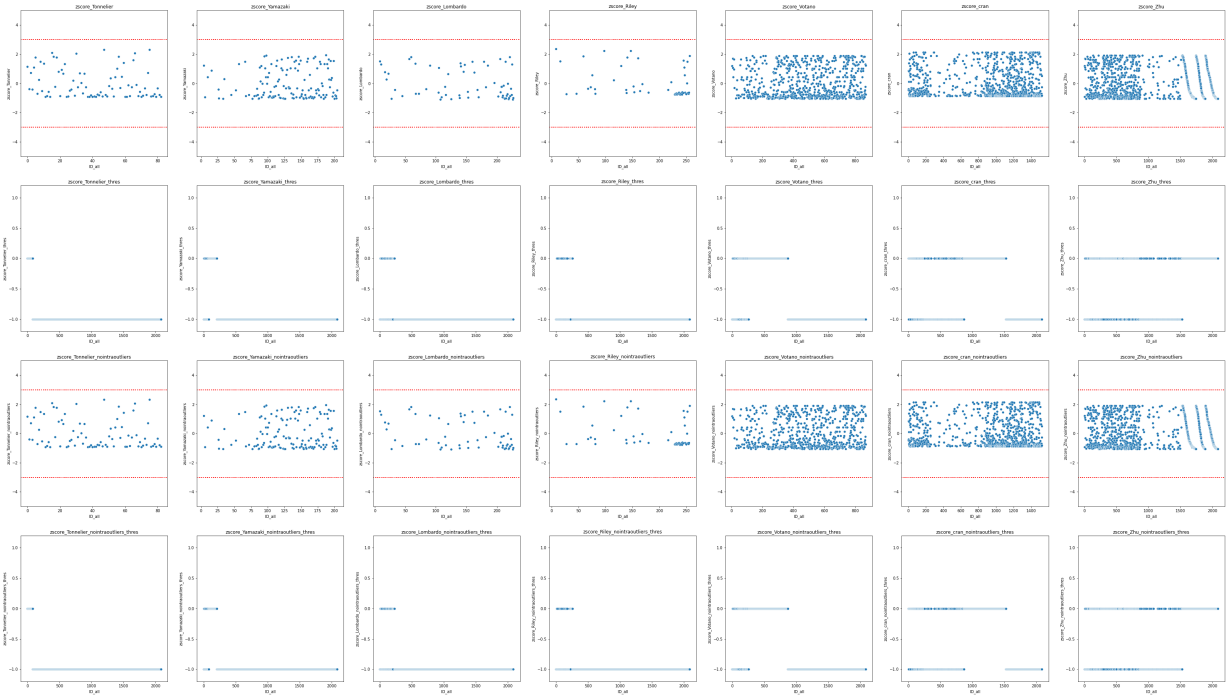


Figure S13: Bioavailability (F30). Inter-dataset Outlier Detection (Categorical Data).

A)



B)

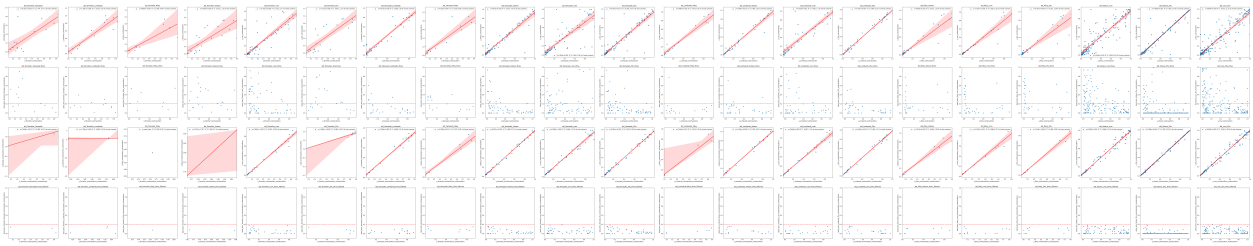


Figure S14: Fraction Unbound (FUB). (A) Intra-dataset Outlier Detection Using Z-scores (Continuous Data); (B) Inter-dataset Outlier Detection Based on Standardized Deviation (Continuous Data).

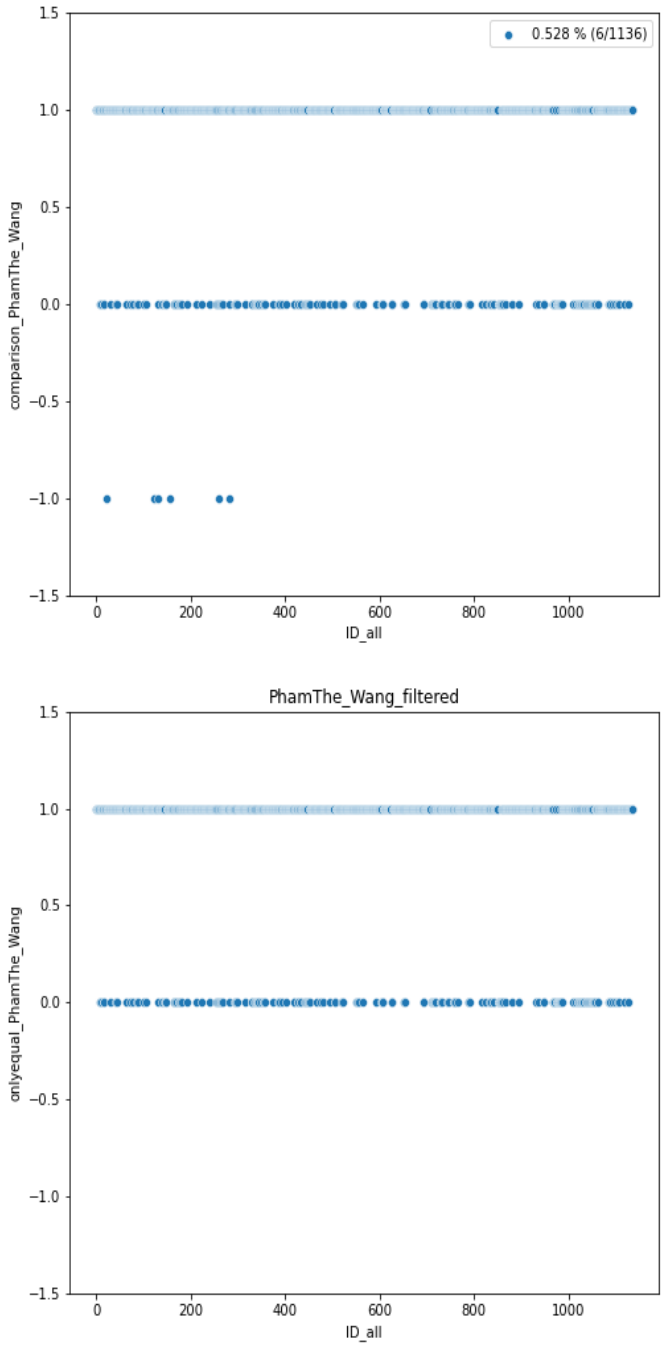


Figure S15: Human Intestinal Absorption (HIA). Inter-dataset Outlier Detection (Categorical Data).

A) **B)**

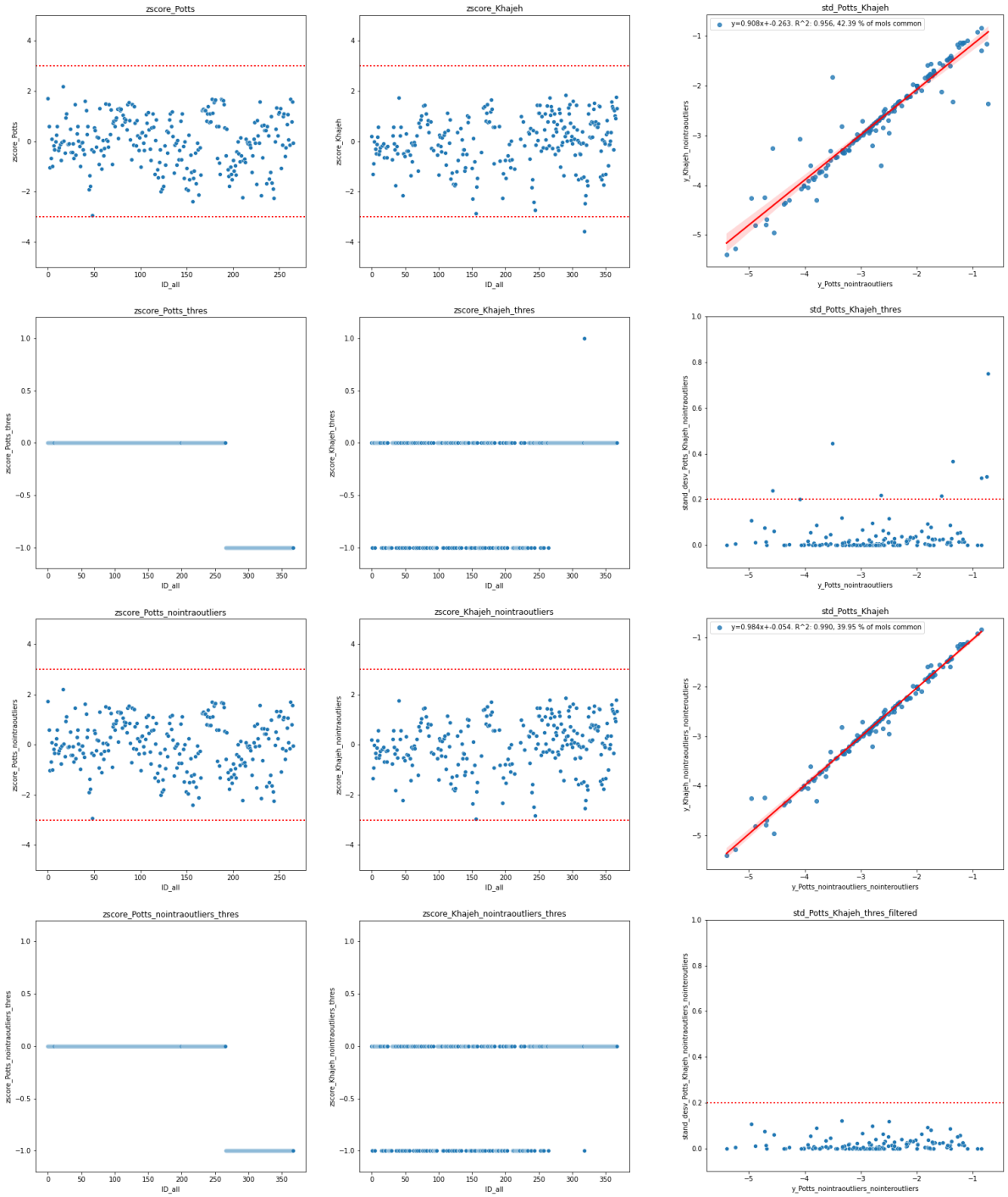


Figure S16: Skin permeation (logKp). (A) Intra-dataset Outlier Detection Using Z-scores (Continuous Data); (B) Inter-dataset Outlier Detection Based on Standardized Deviation (Continuous Data).

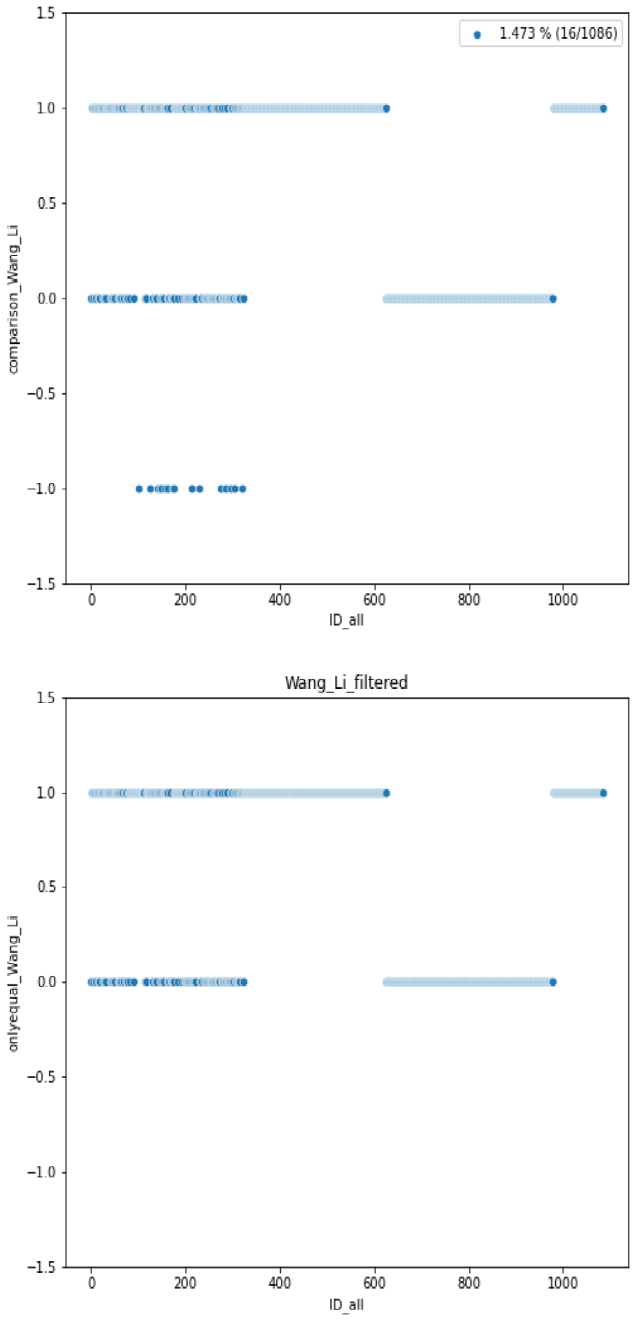


Figure S17: P-glycoprotein substrate (Pgp.sub). Inter-dataset Outlier Detection (Categorical Data).