

Supplementary figures

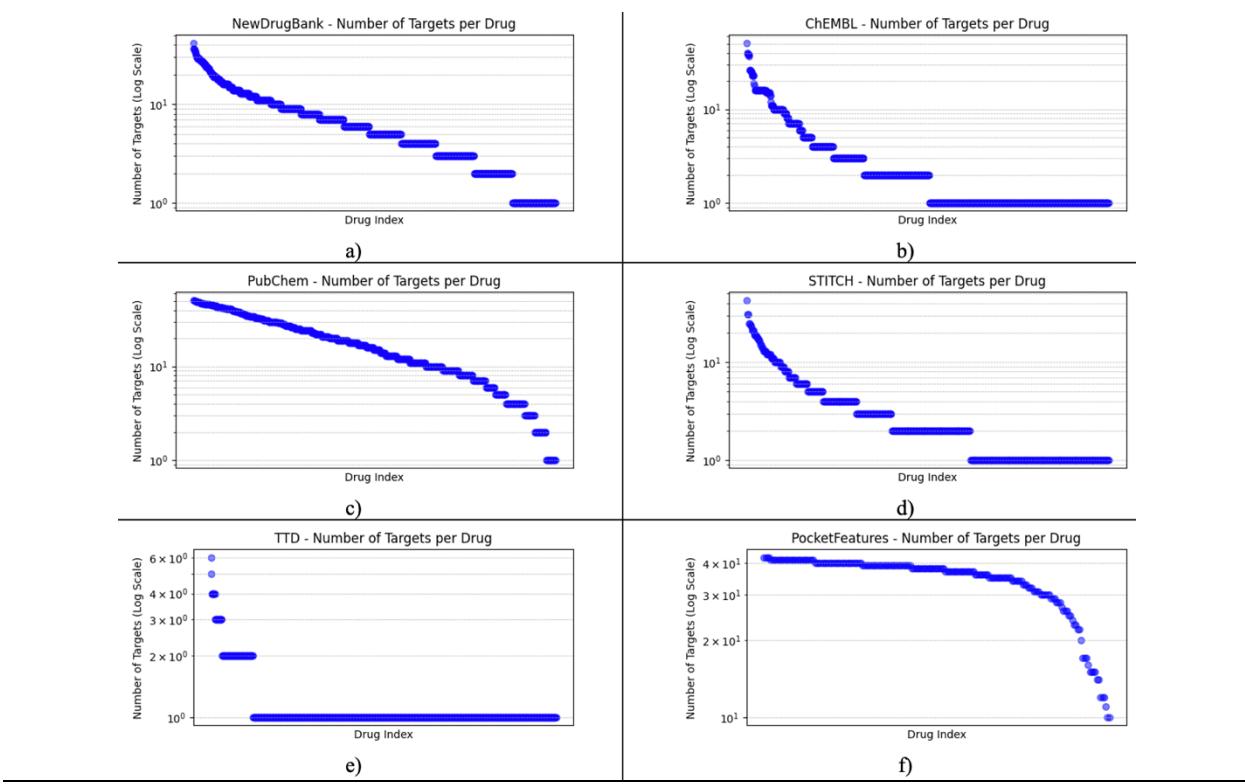


Figure S1. Scatter plots showing the number of targets per drug across six databases. The logarithmic scale reveals that most drugs interact with a few targets, indicating high selectivity, with a rapid decline in target numbers per drug.

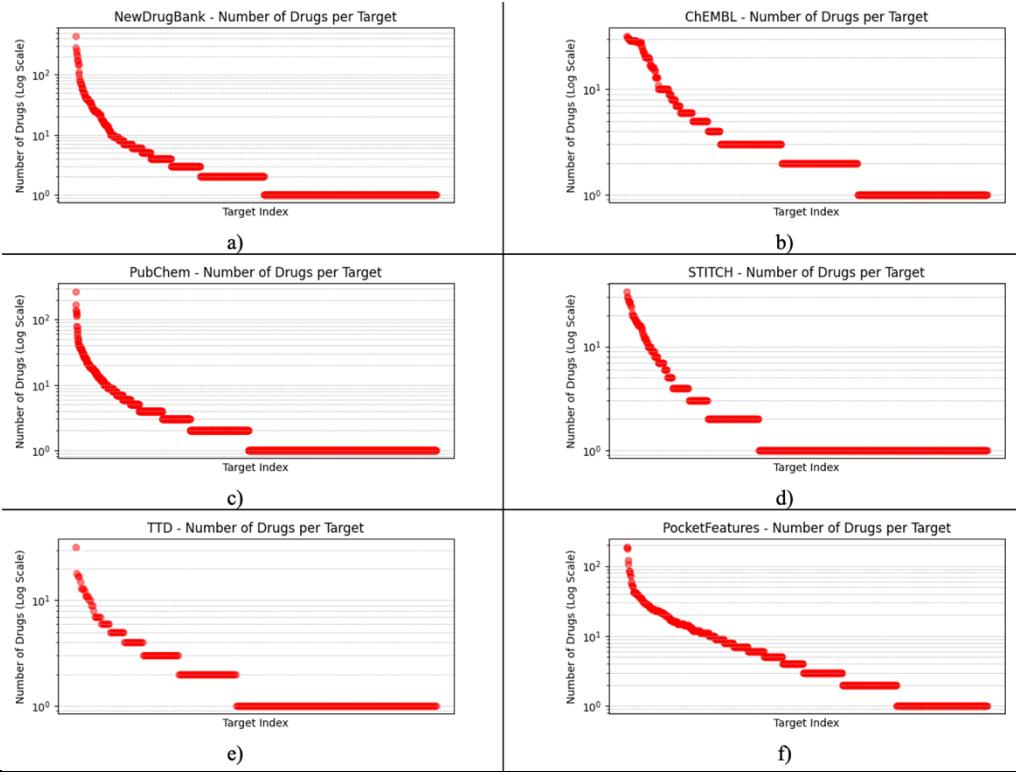
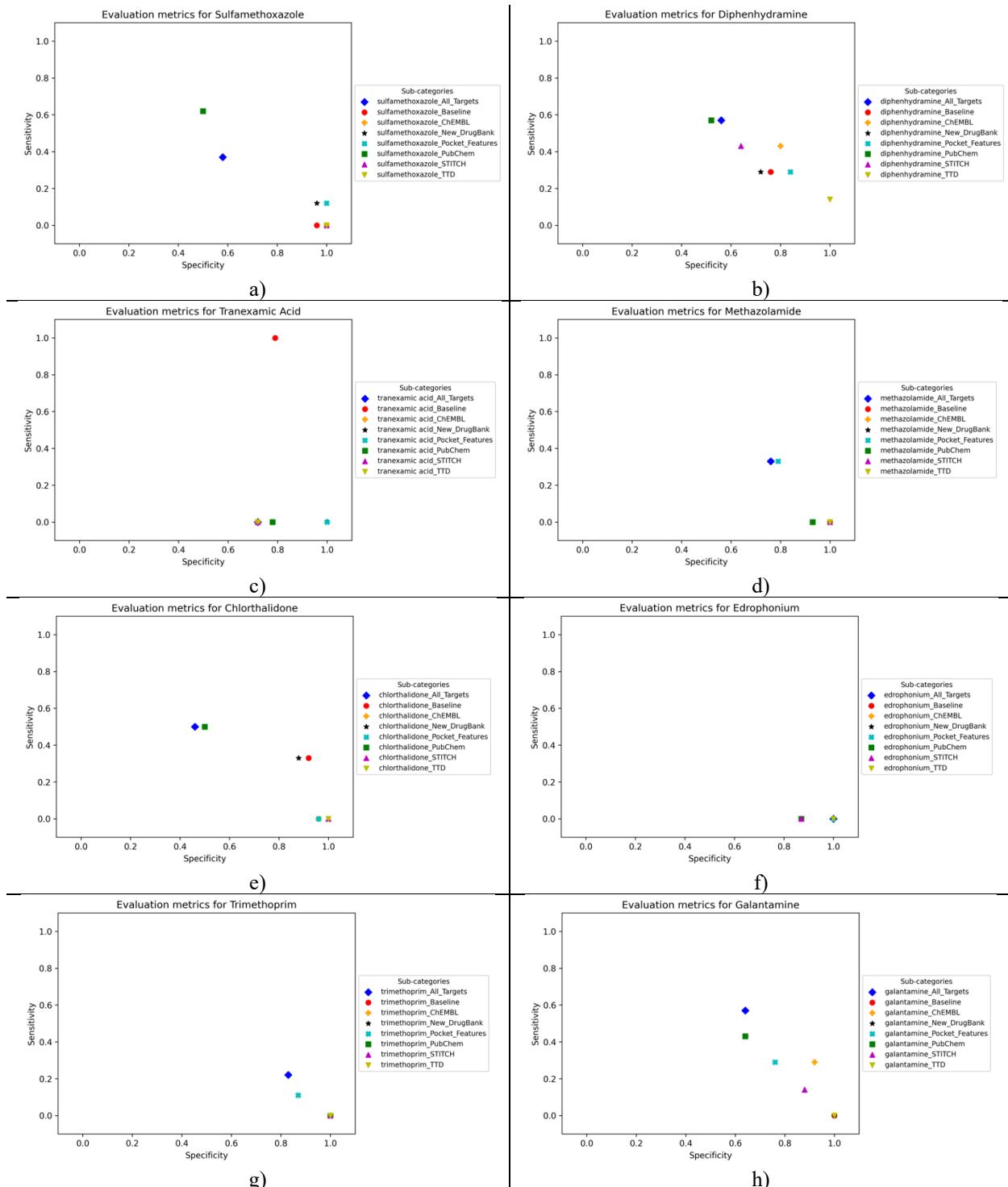
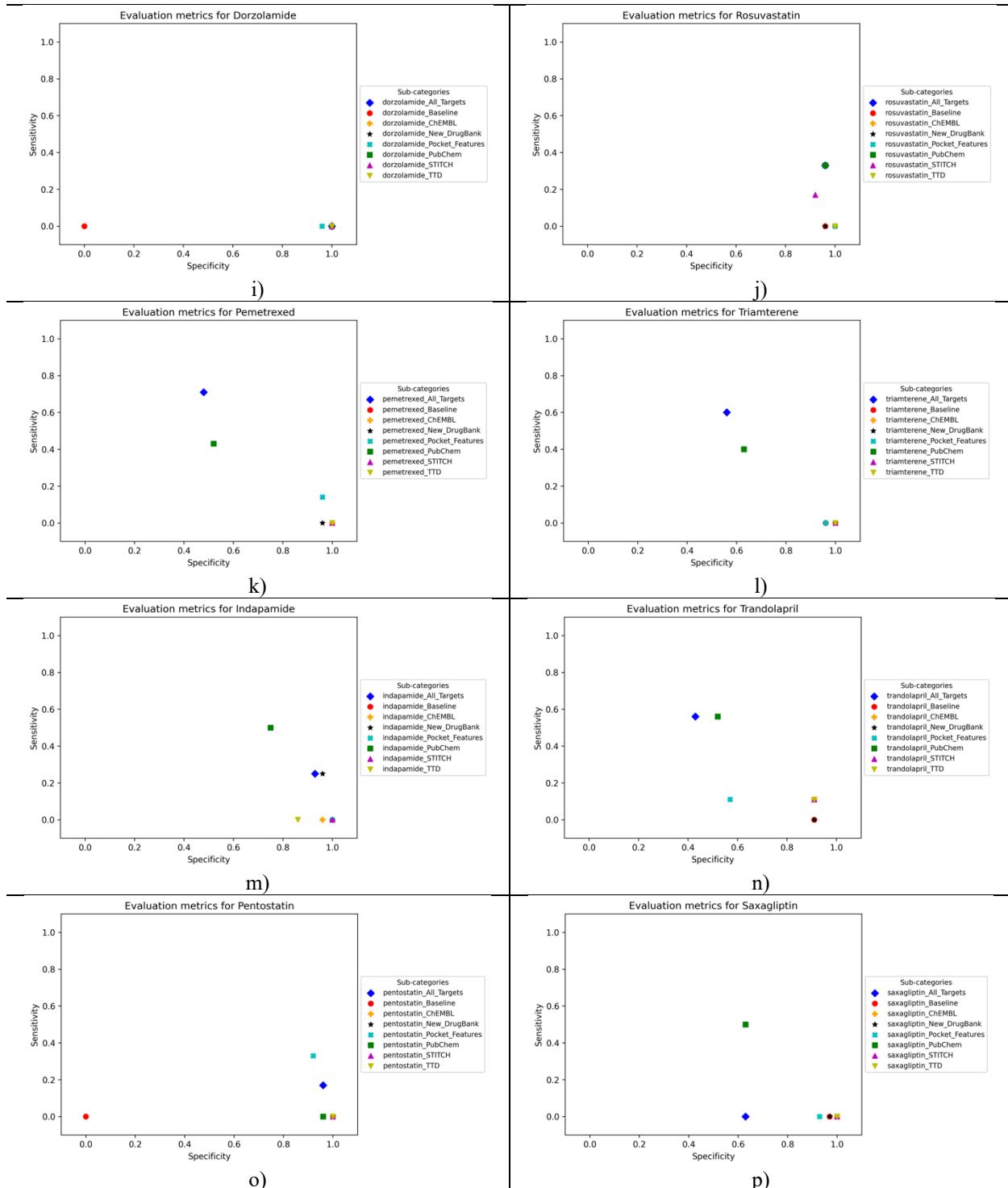
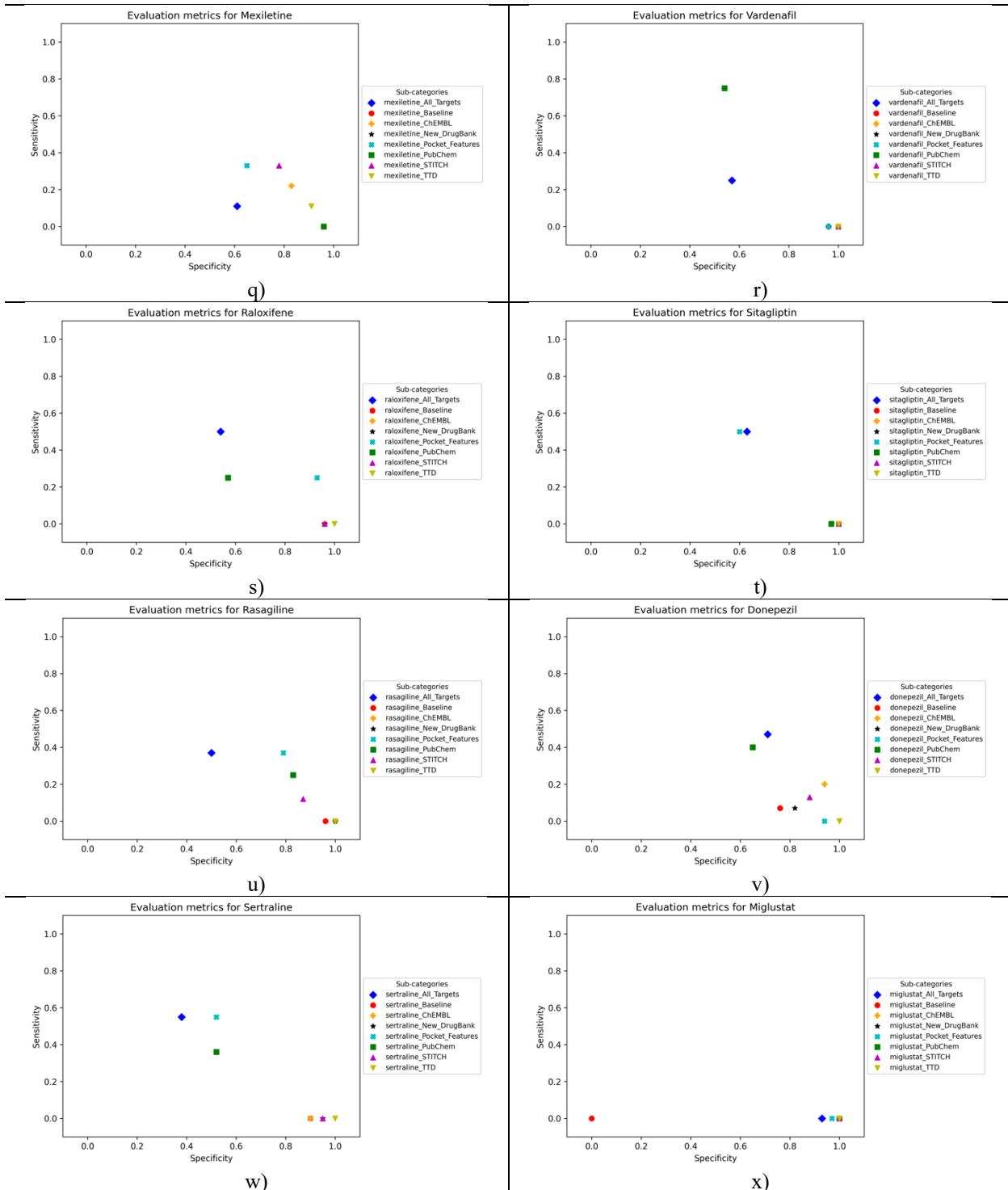


Figure S2. Scatter plots depicting the number of drugs per target across six databases. The logarithmic scale highlights that a few targets interact with many drugs, while most targets interact with fewer drugs, demonstrating target promiscuity and specialization.







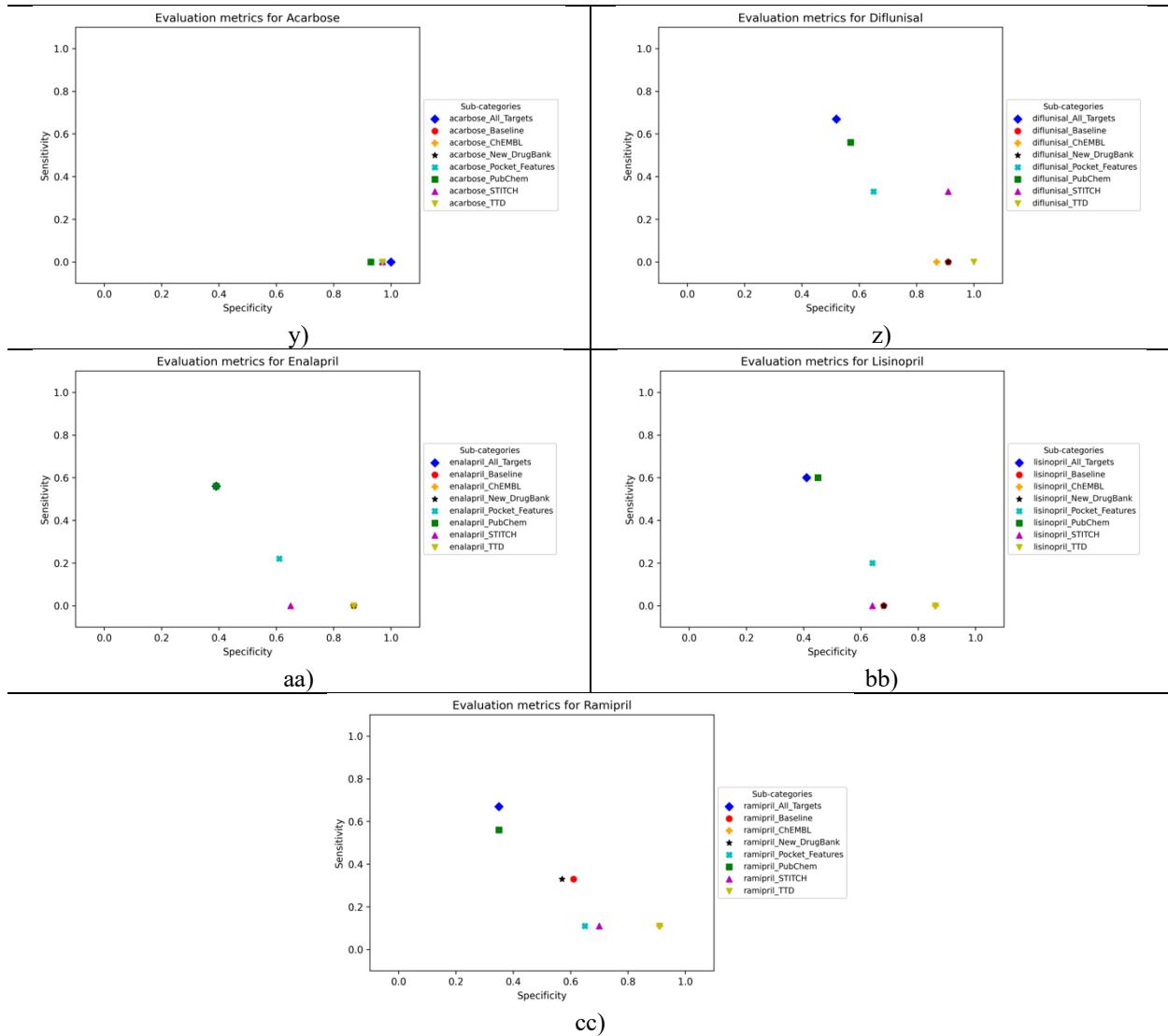
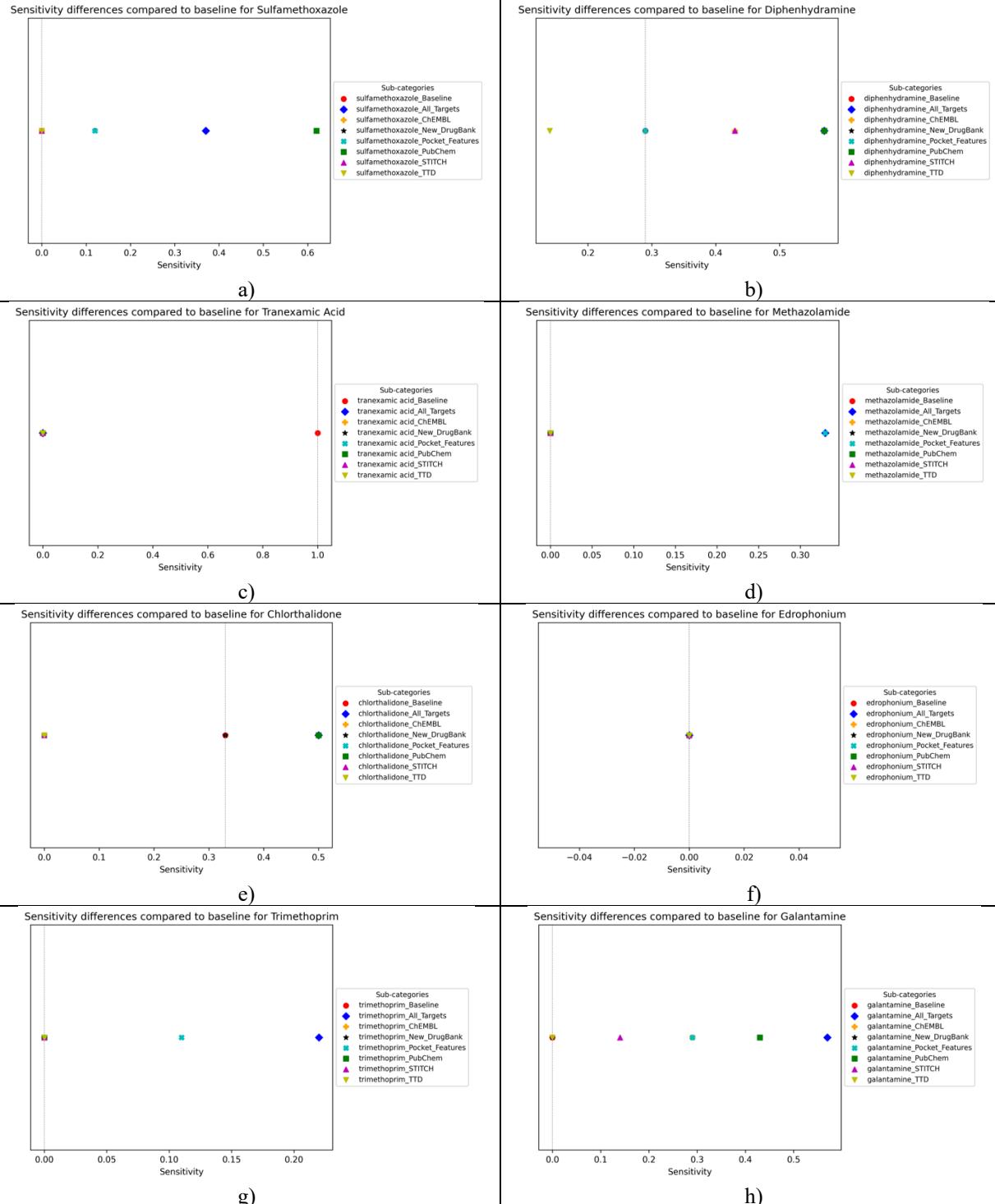
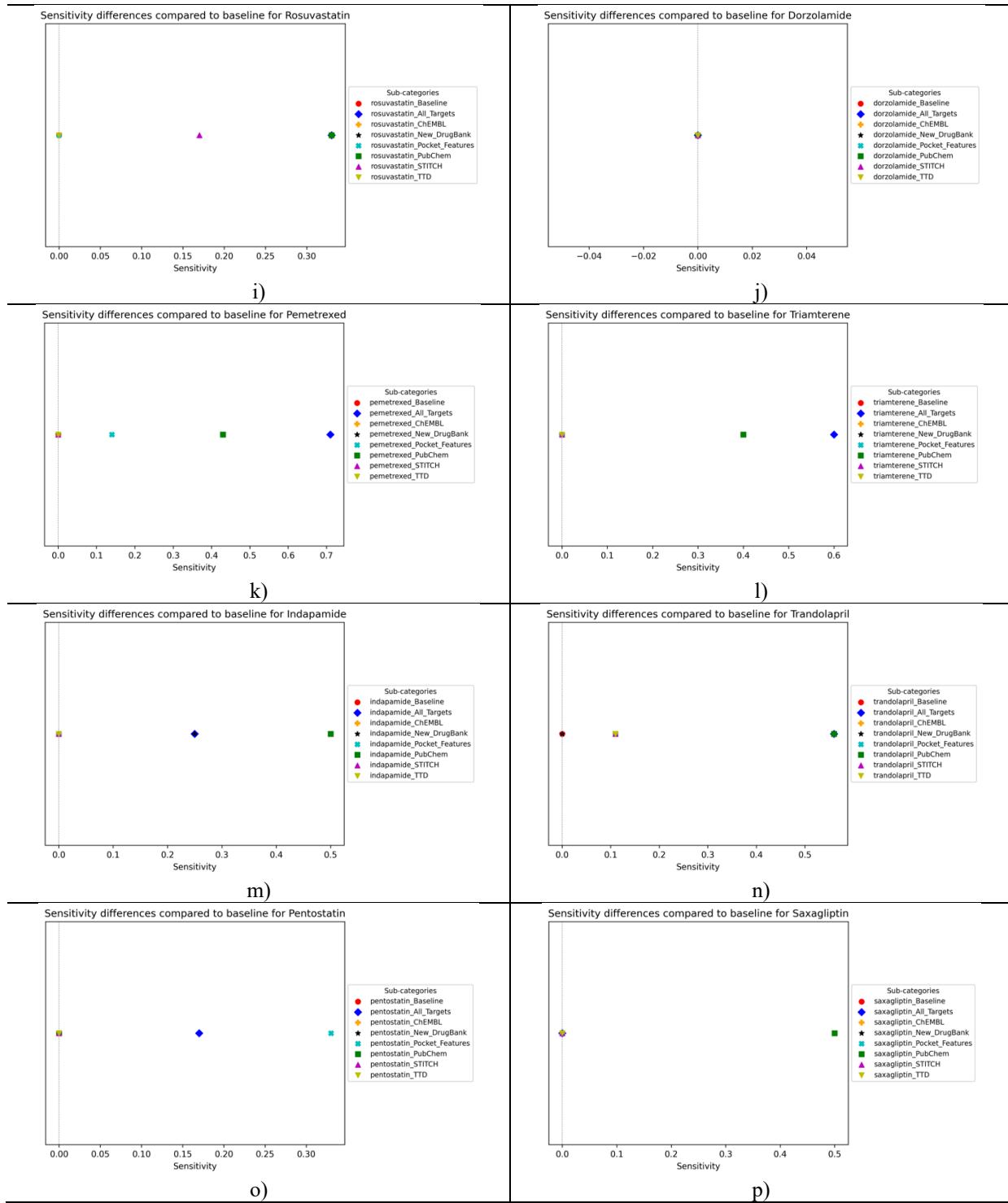
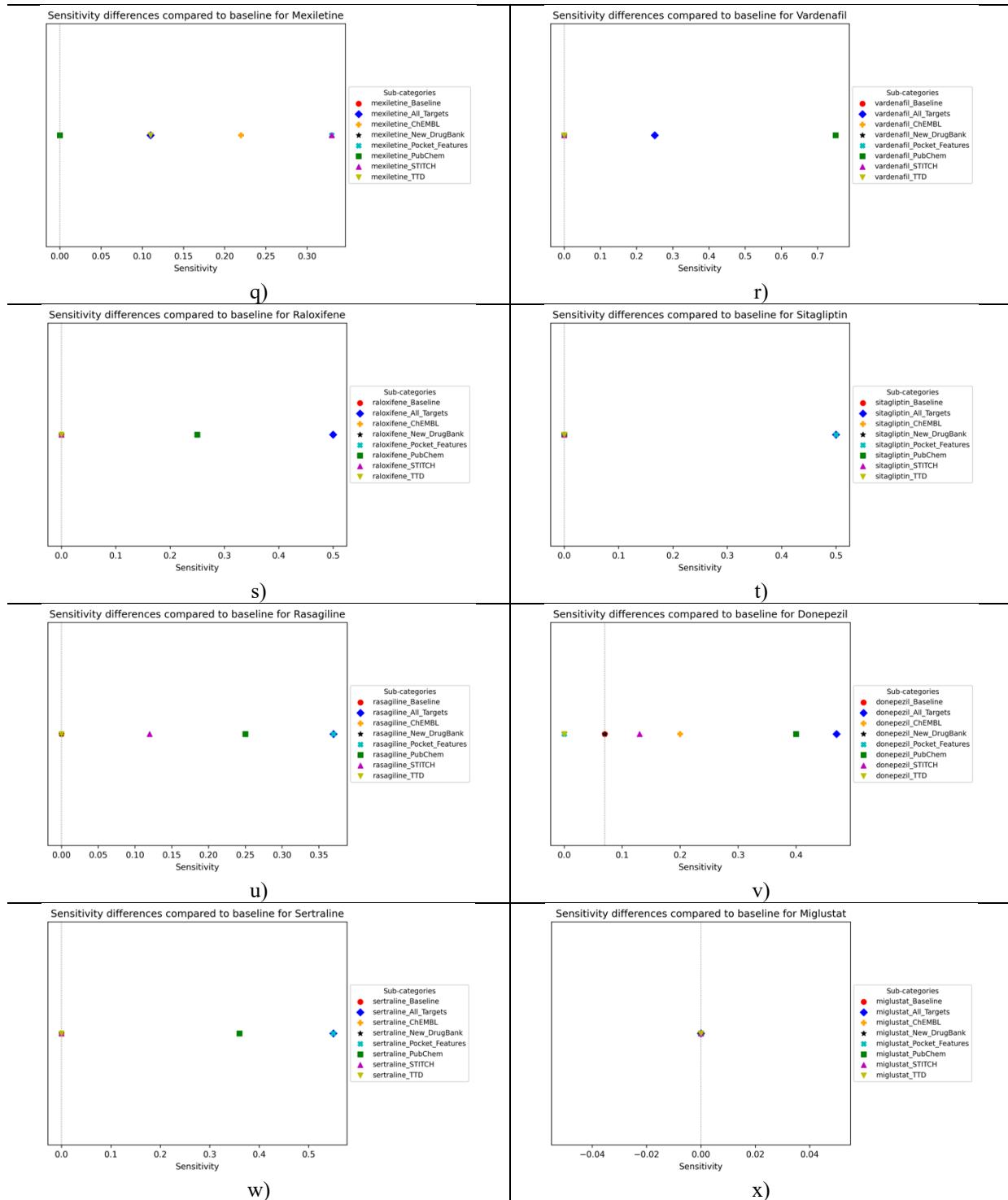


Figure S3. Evaluation (sensitivity-specificity) plots for all selected drugs across all selected databases and the baseline.







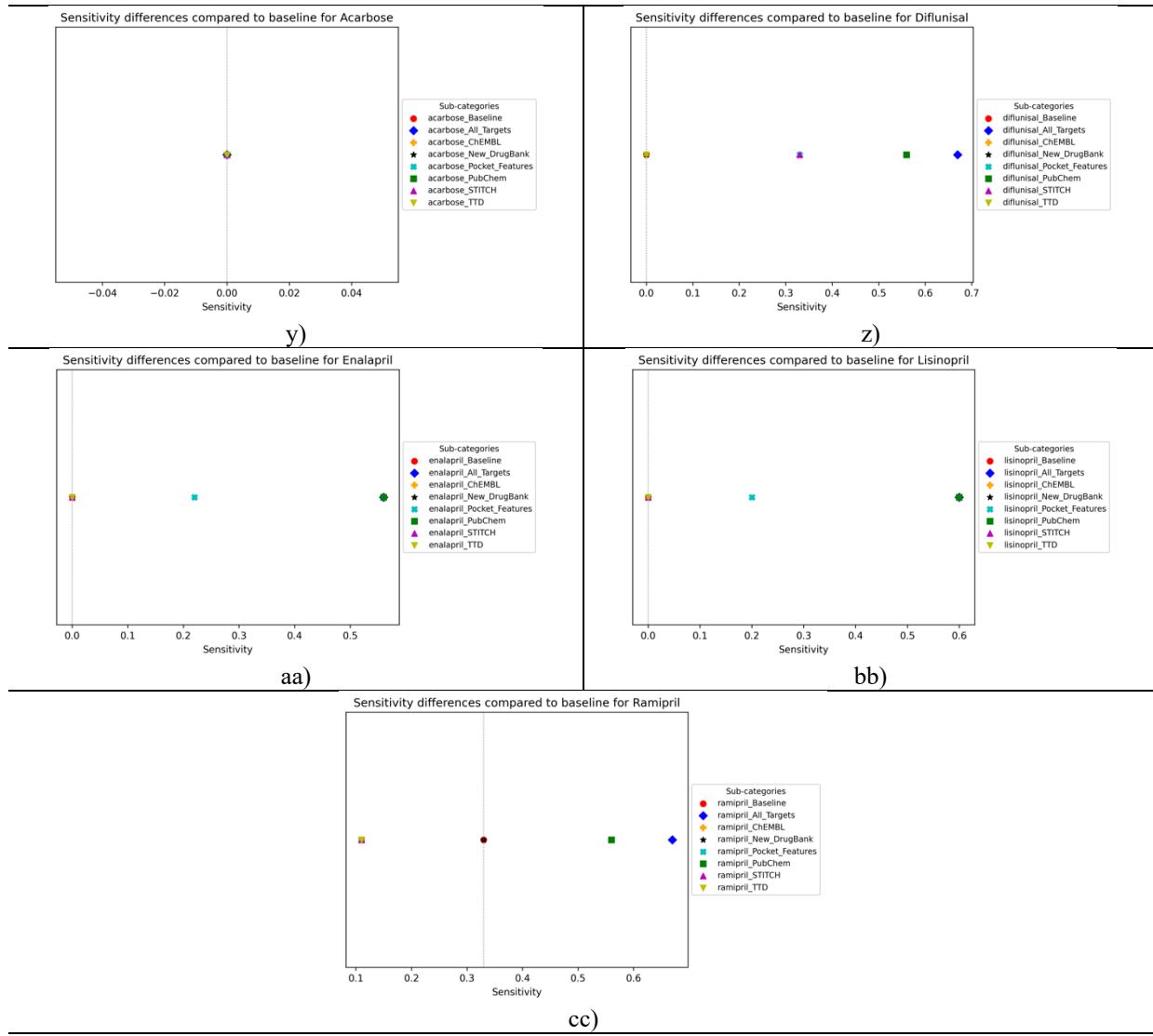
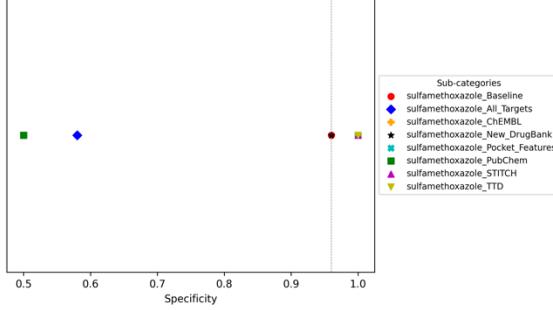


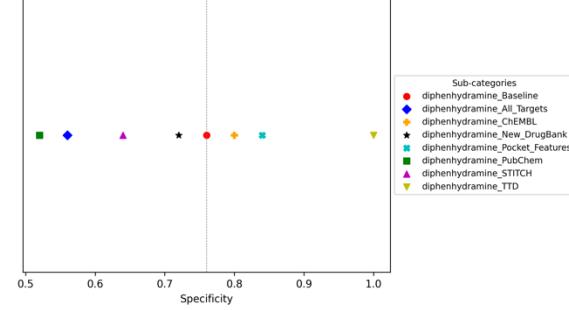
Figure S4. Comparisons of sensitivities for all selected drugs across all selected databases and the baseline.

Specificity differences compared to baseline for Sulfamethoxazole



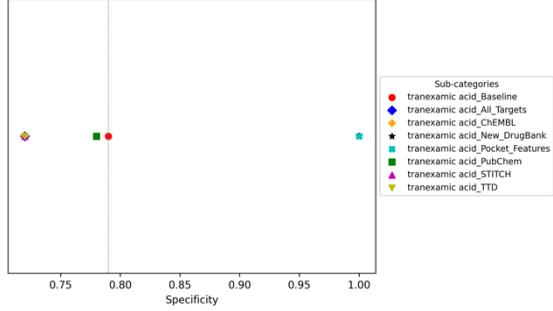
a)

Specificity differences compared to baseline for Diphenhydramine



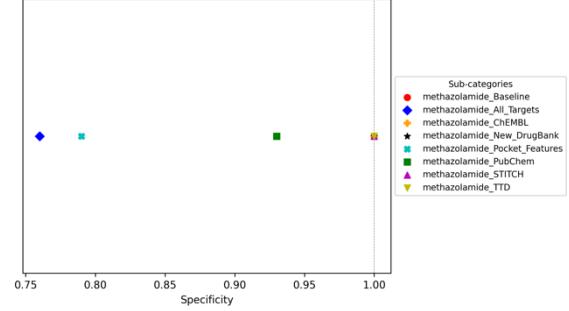
b)

Specificity differences compared to baseline for Tranexamic Acid



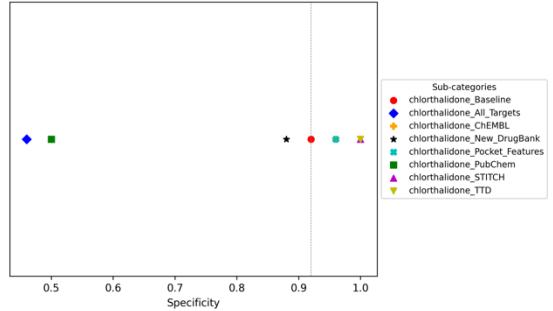
c)

Specificity differences compared to baseline for Methazolamide



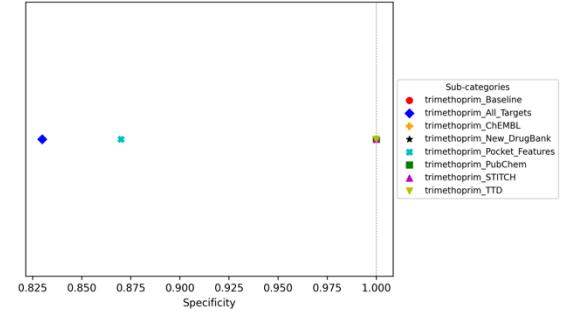
d)

Specificity differences compared to baseline for Chlorthalidone



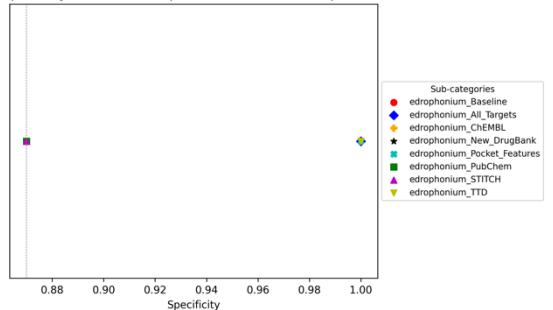
e)

Specificity differences compared to baseline for Trimethoprim



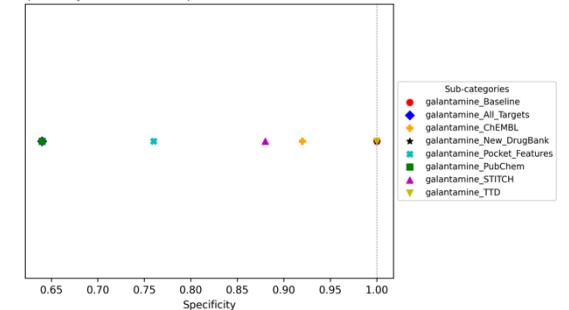
f)

Specificity differences compared to baseline for Edrophonium

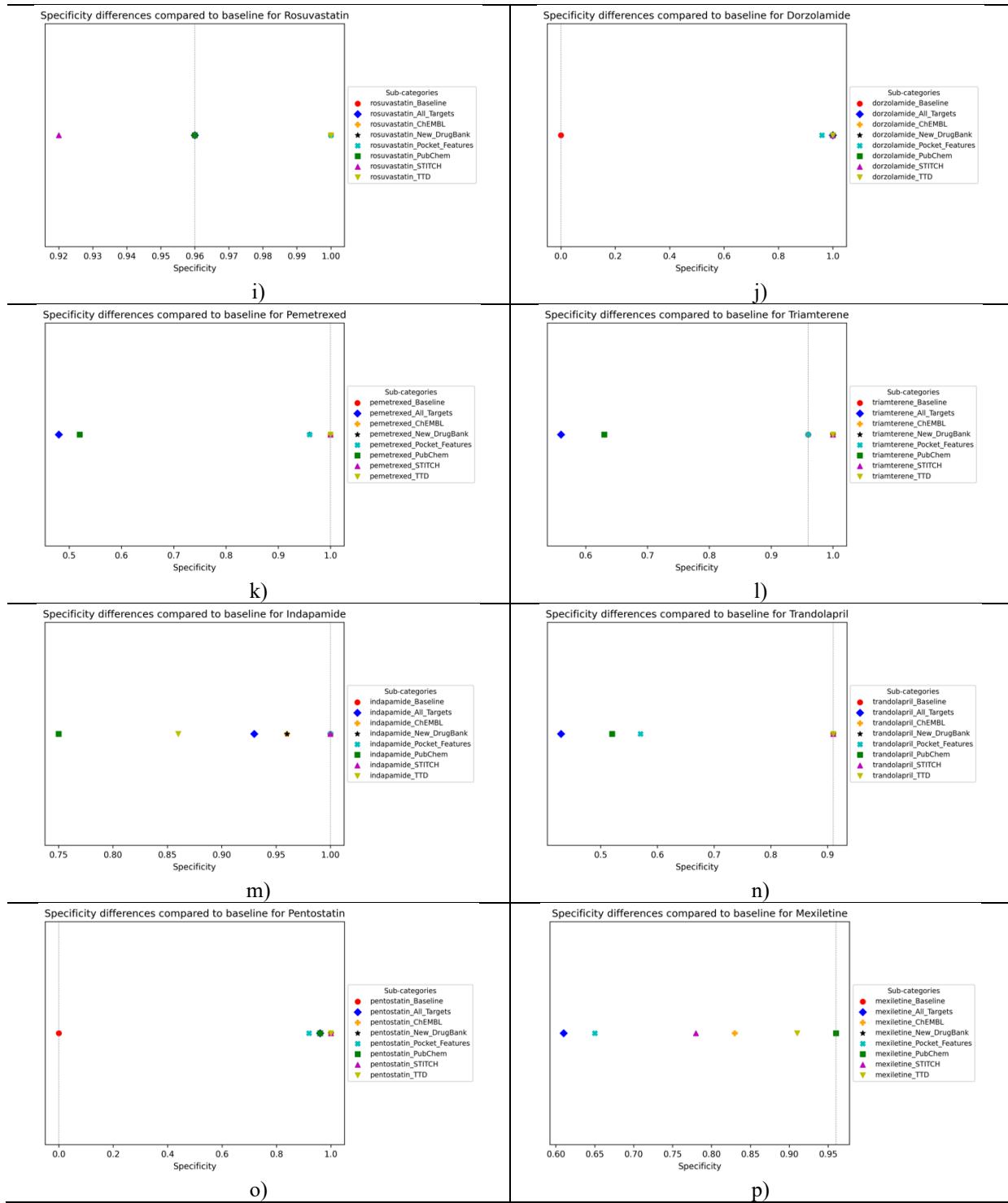


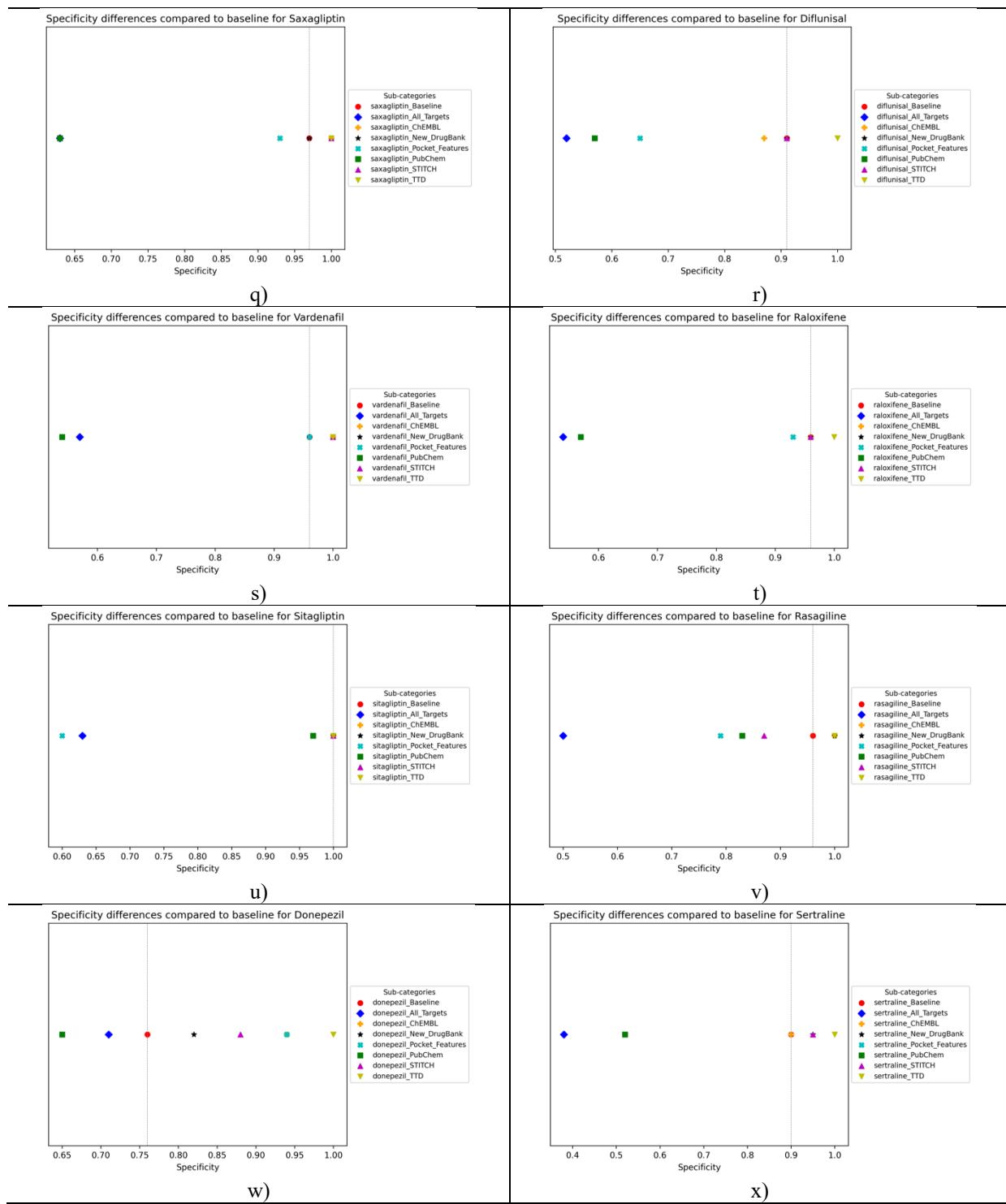
g)

Specificity differences compared to baseline for Galantamine



h)





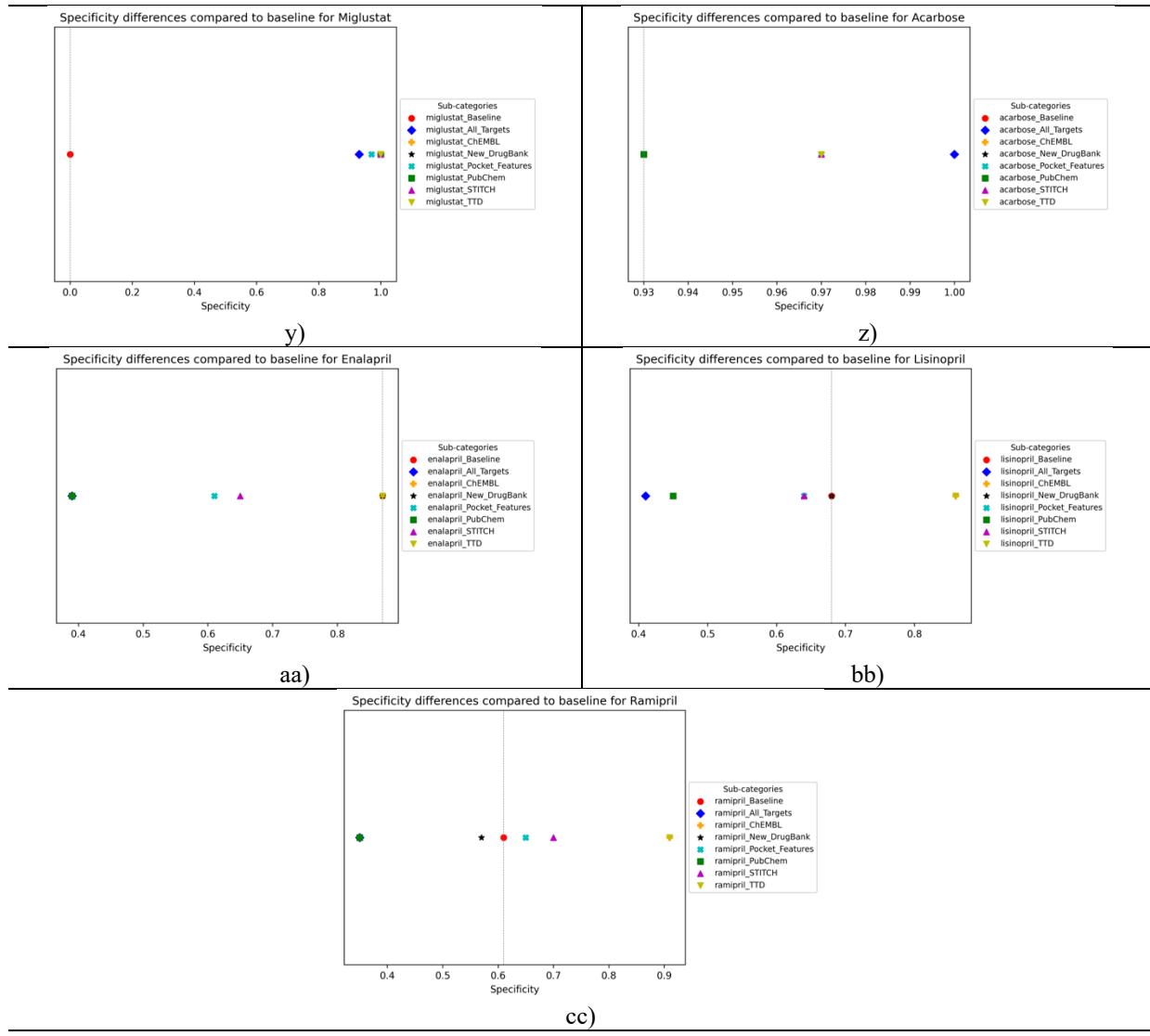
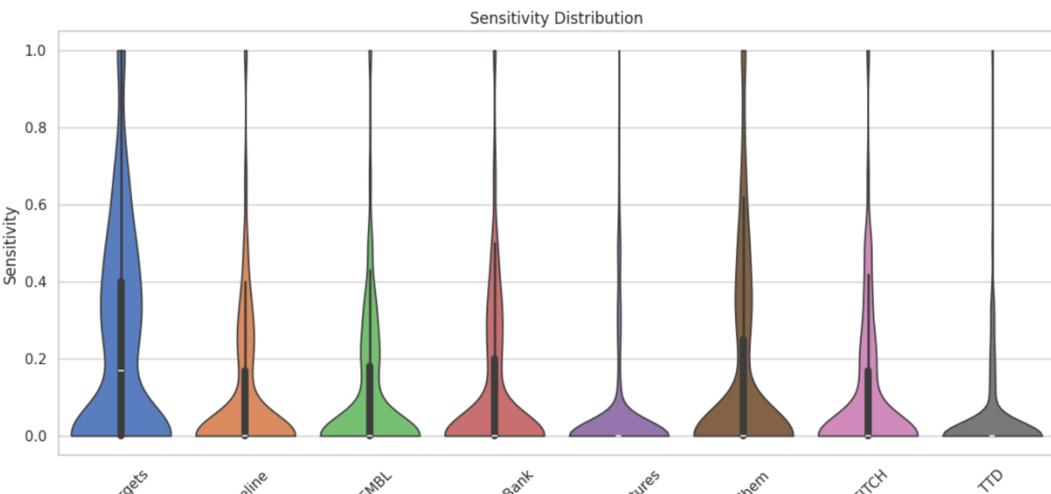
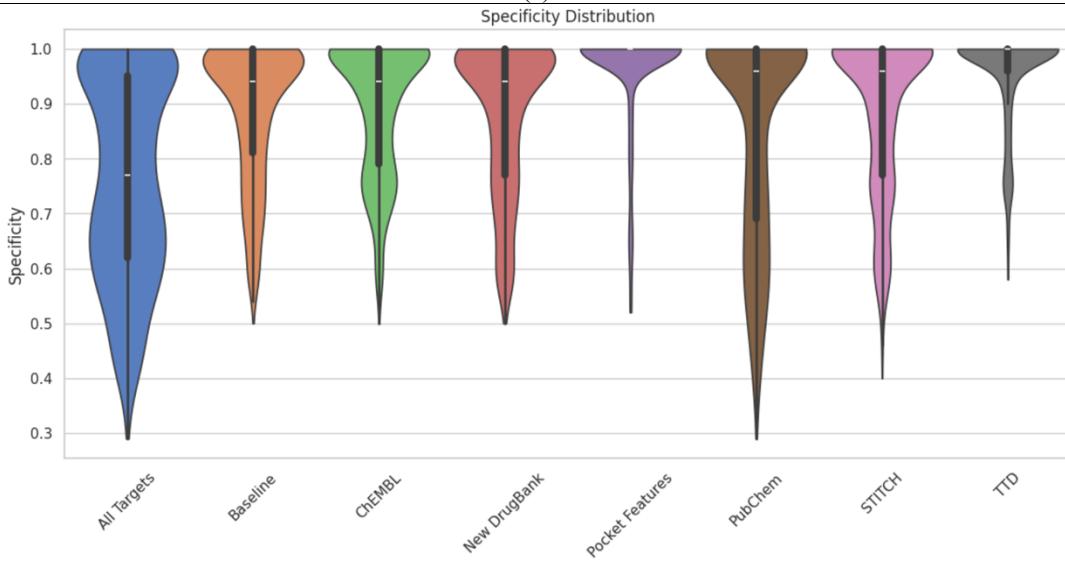


Figure S5. Comparisons of specificities for all selected drugs across all selected databases and the baseline.

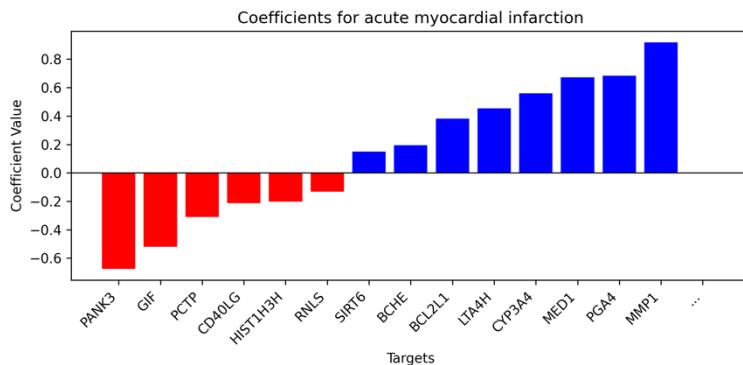


(a)

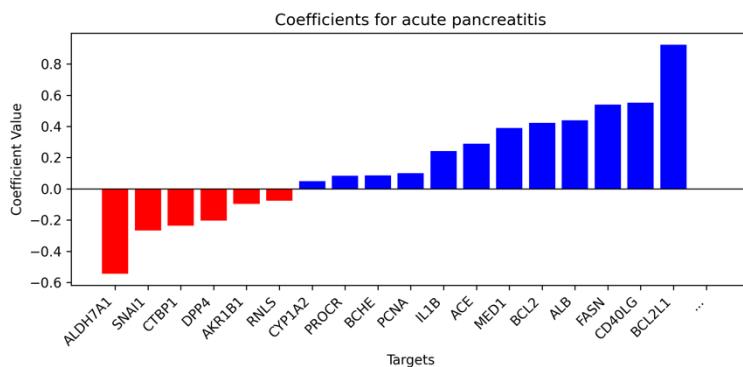


(b)

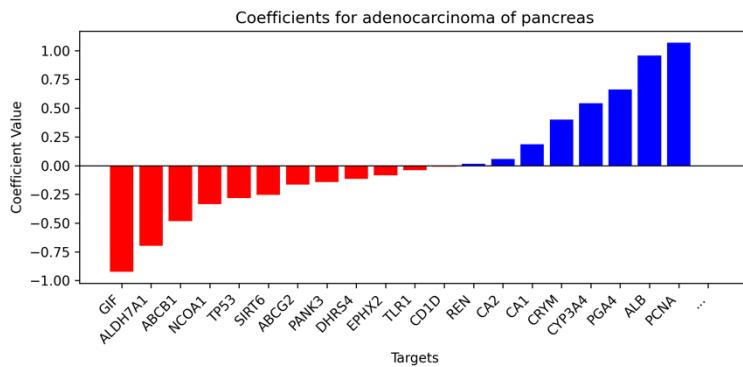
Figure S6. Violin plots illustrating the distributions of (a) sensitivity and (b) specificity for 565 drugs present in at least four databases. Sensitivity improves when combining targets from all databases, as reflected in the wider range of high-sensitivity values. Specificity remains high for individual databases but decreases when integrating all targets, demonstrating the trade-off between these metrics.



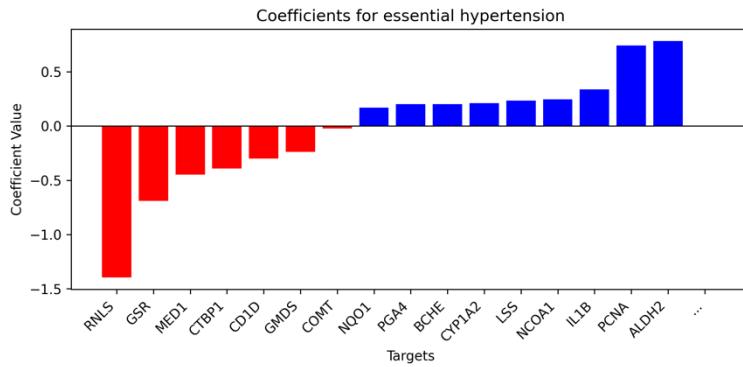
a)



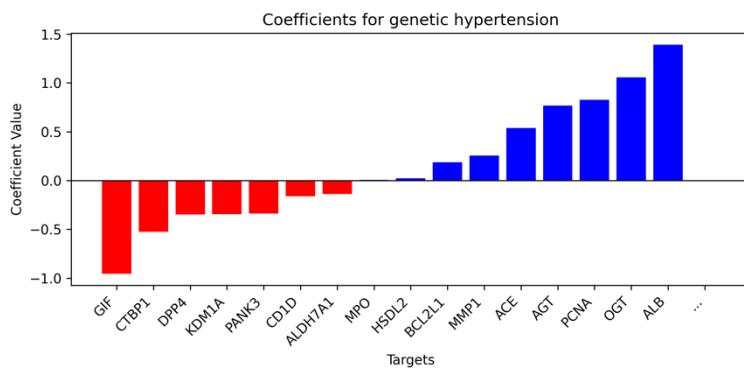
b)



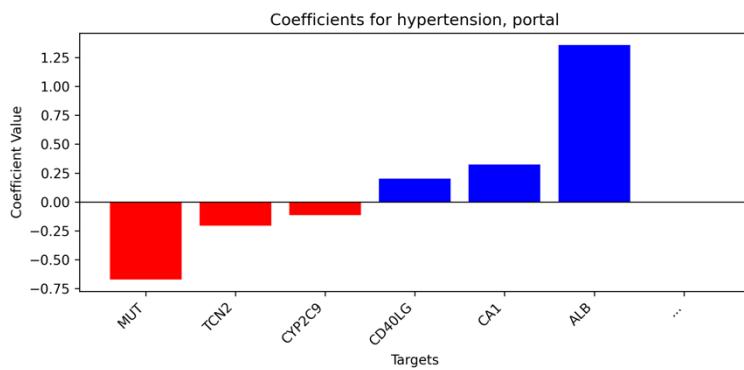
c)



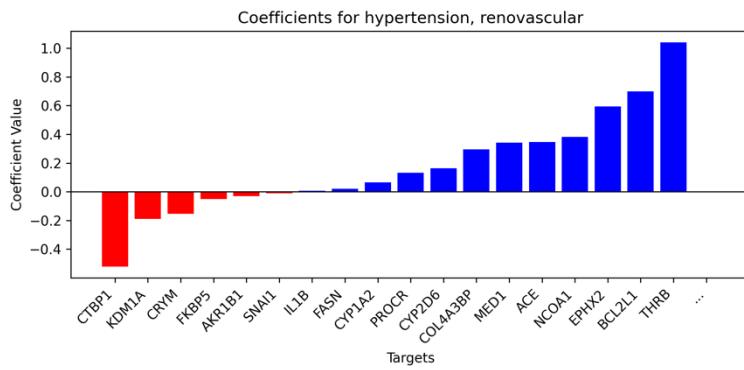
d)



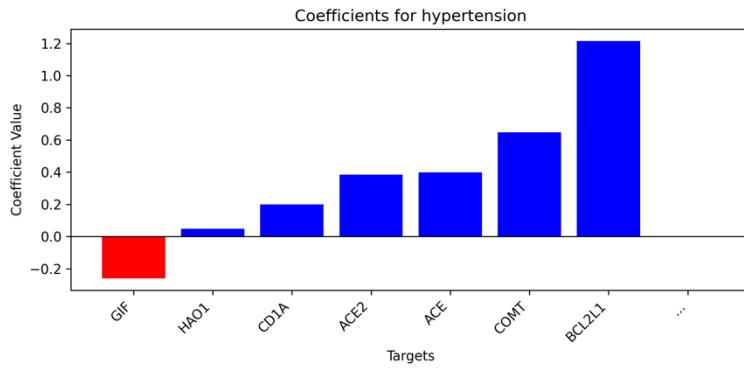
e)



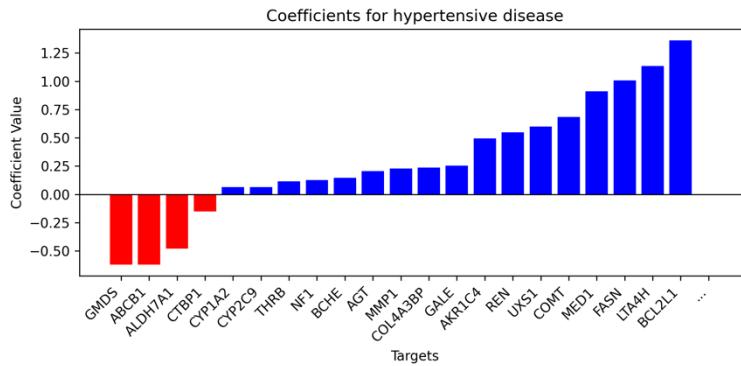
f)



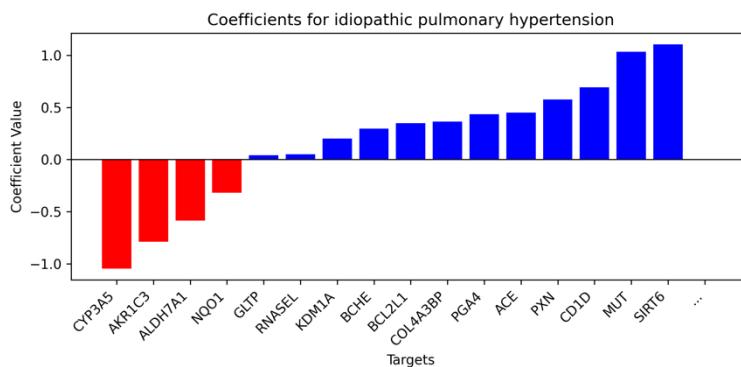
g)



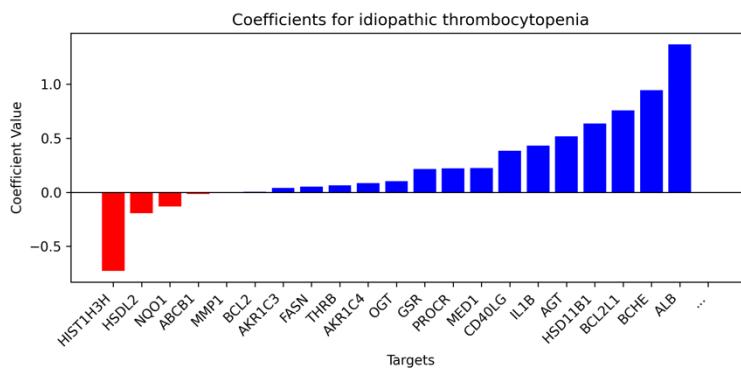
h)



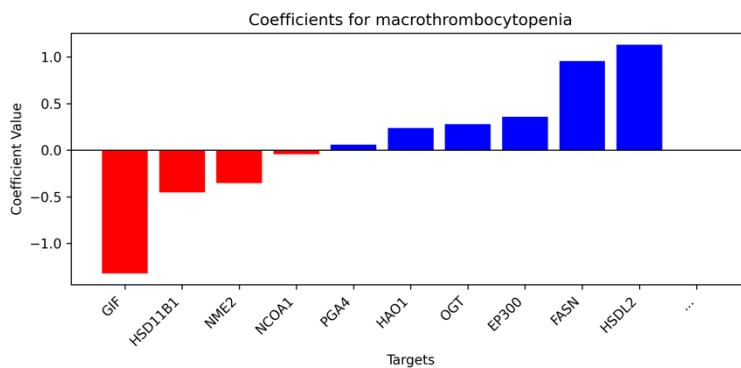
i)



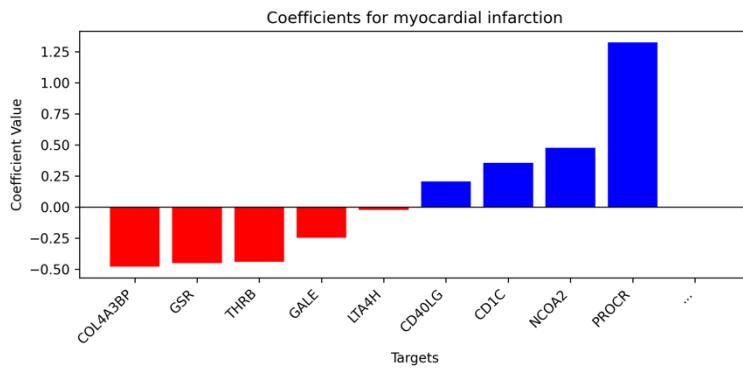
j)



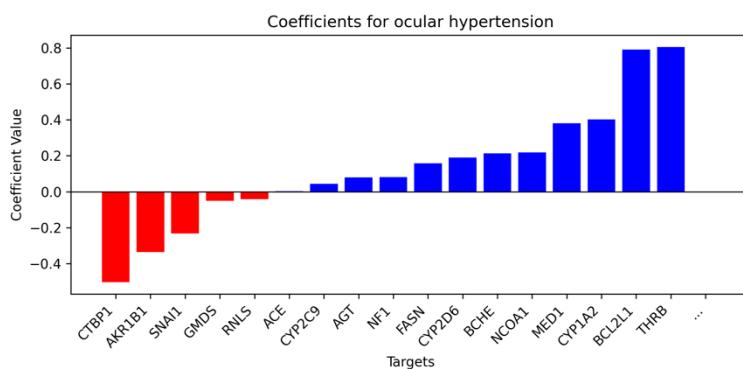
k)



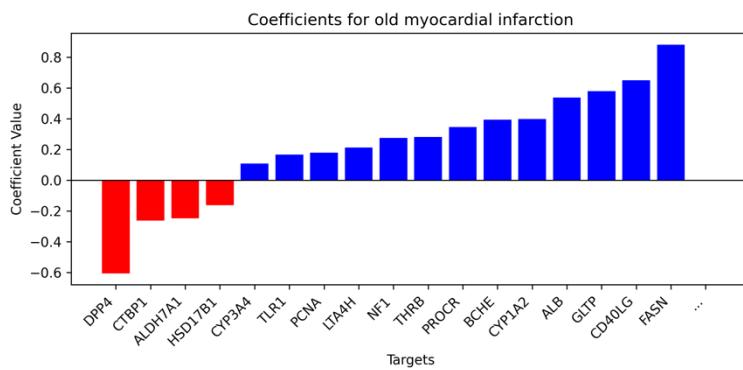
l)



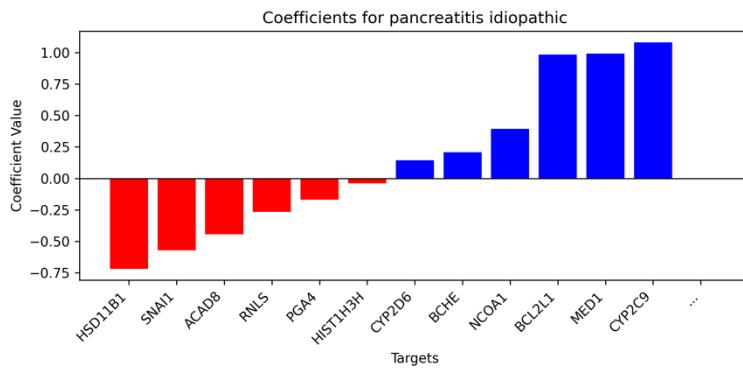
m)



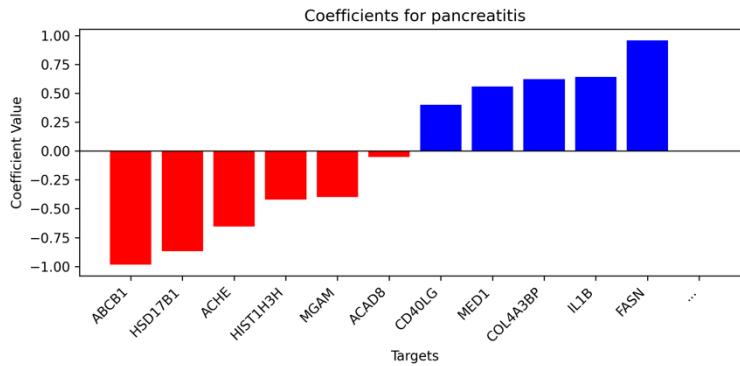
n)



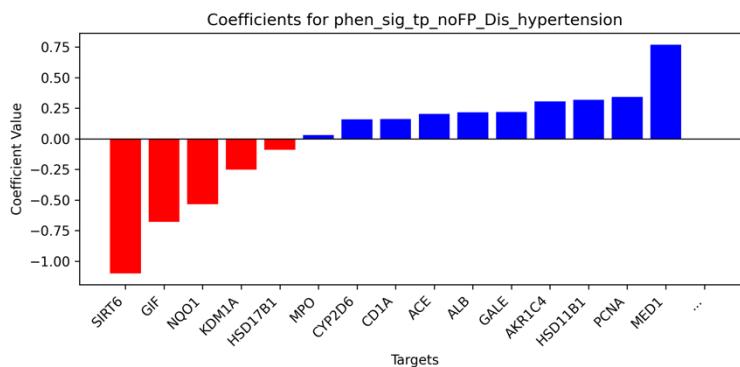
o)



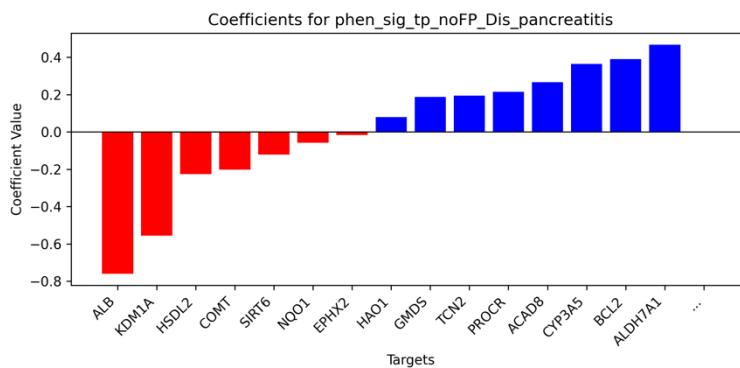
p)



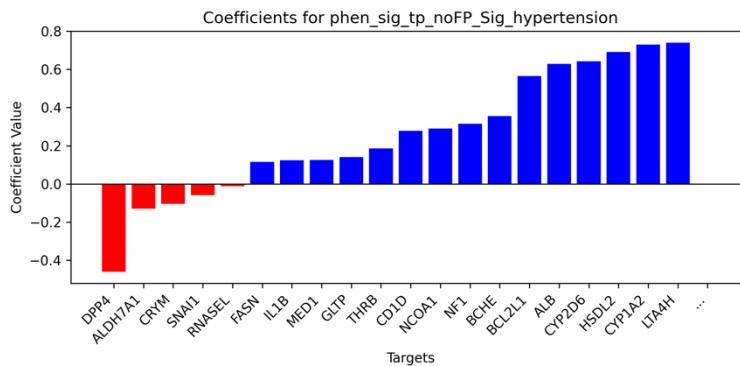
q)



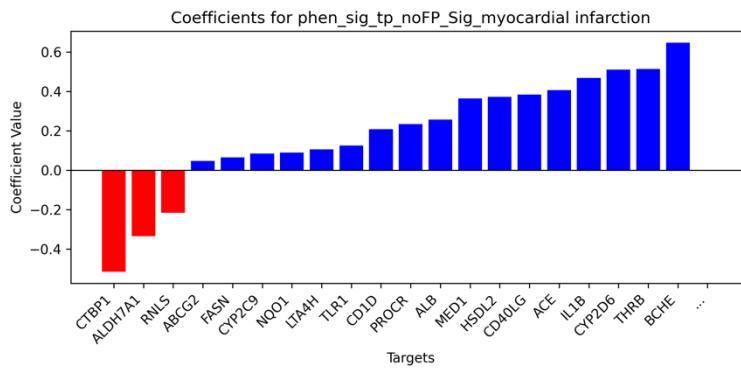
r)



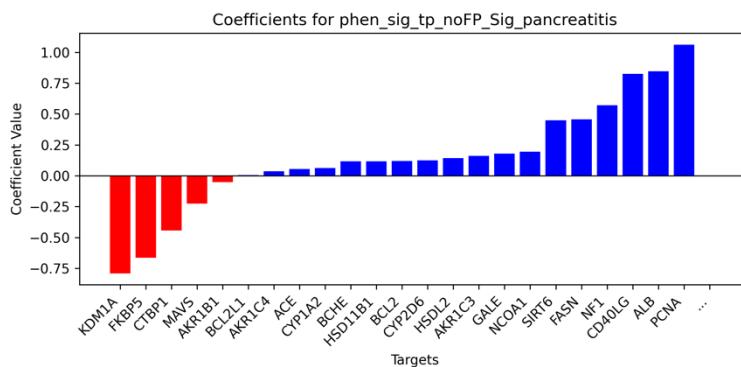
s)



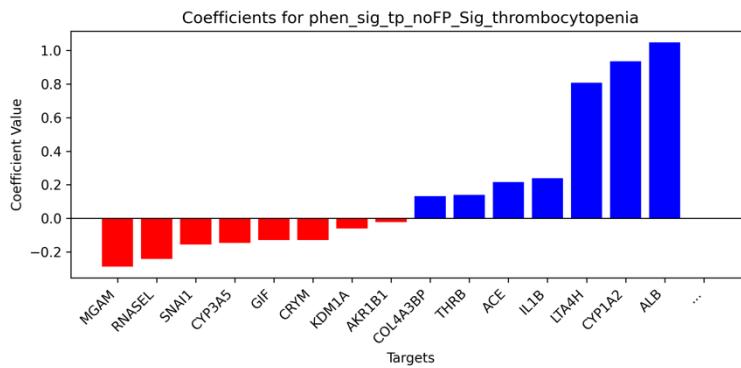
t)



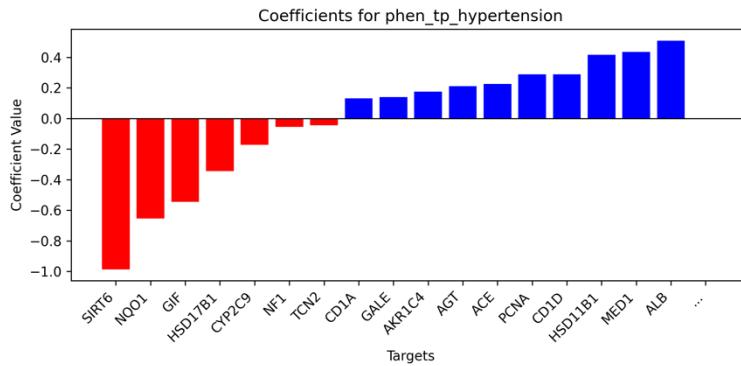
u)



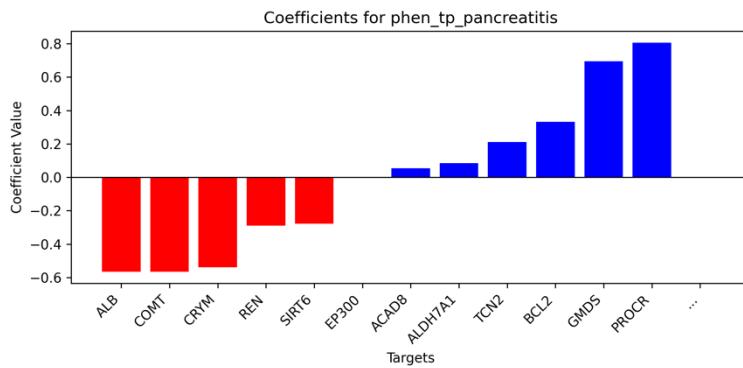
v)



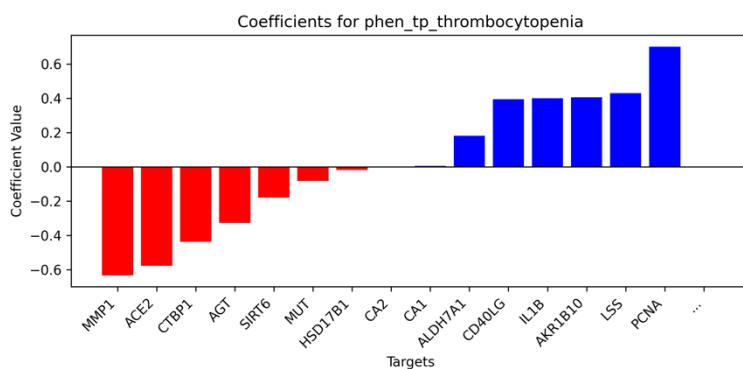
w)



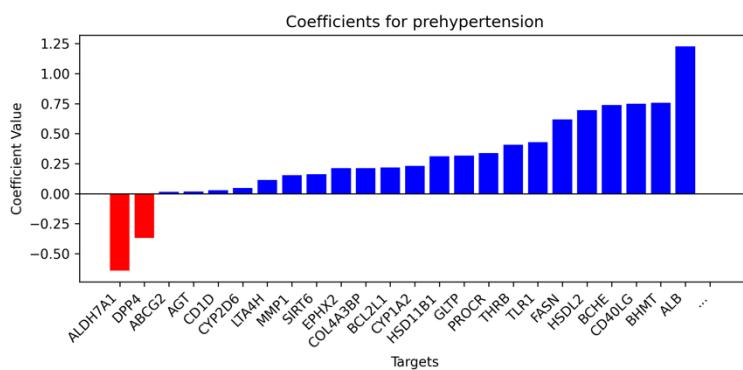
x)



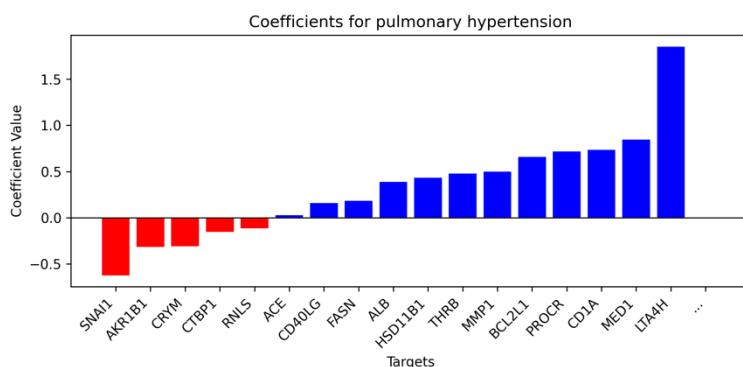
y)



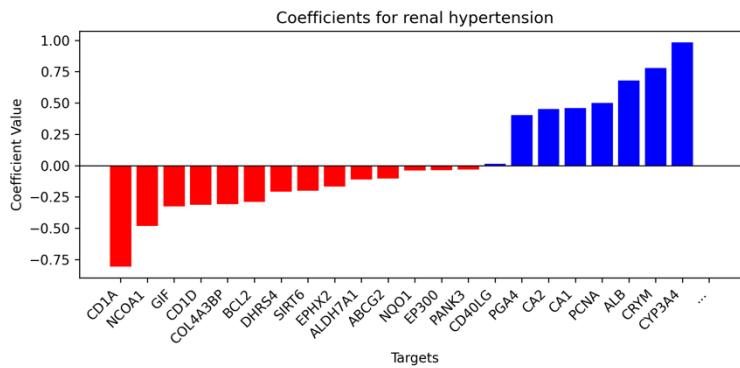
z)



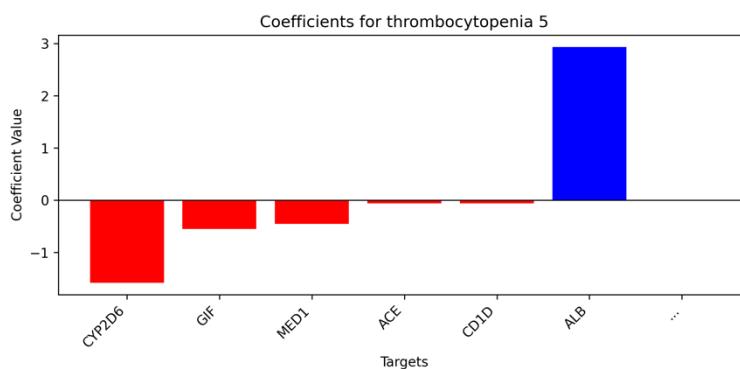
aa)



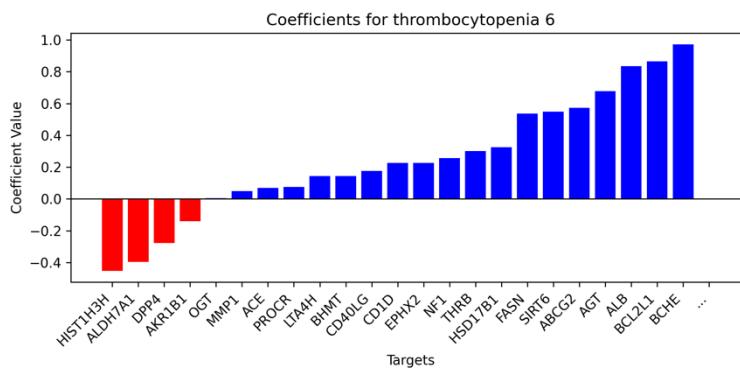
bb)



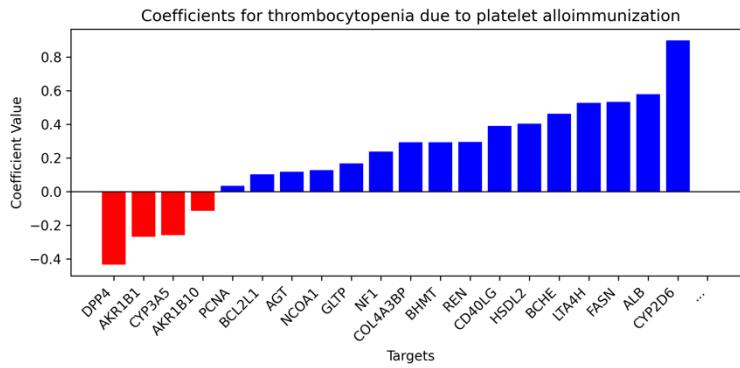
cc)



dd)



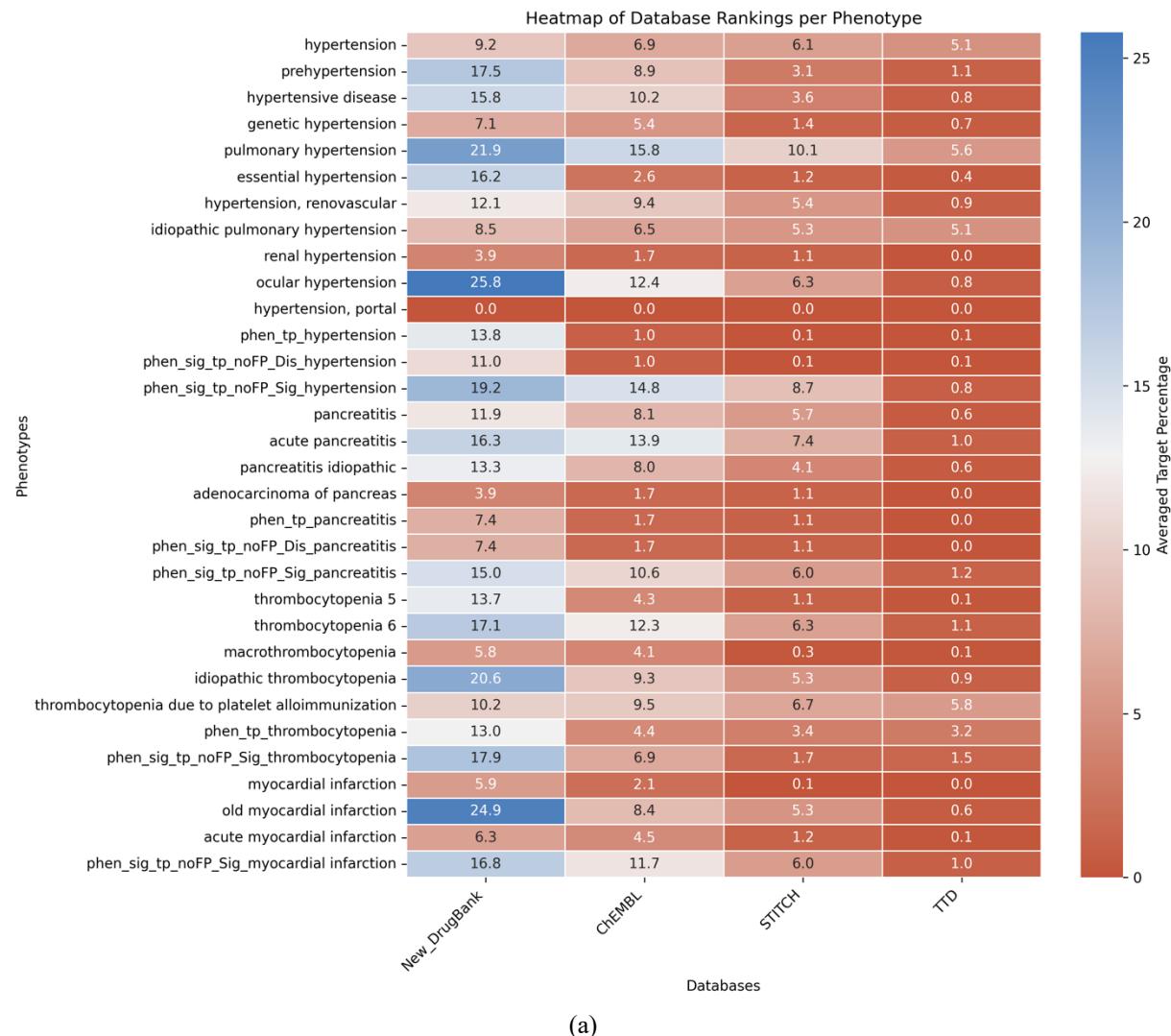
ee)



ff)

Figure S7. The logistic regression coefficient plots for targets associated with all 32 phenotypes in the analysis. Positive coefficients (blue bars) indicate targets that have a positive association with prehypertension, while

negative coefficients (red bars) indicate targets with an inverse relationship. The ellipsis (“...”) at the end of the x-axis signifies the presence of additional targets with zero coefficients that are not displayed for readability.



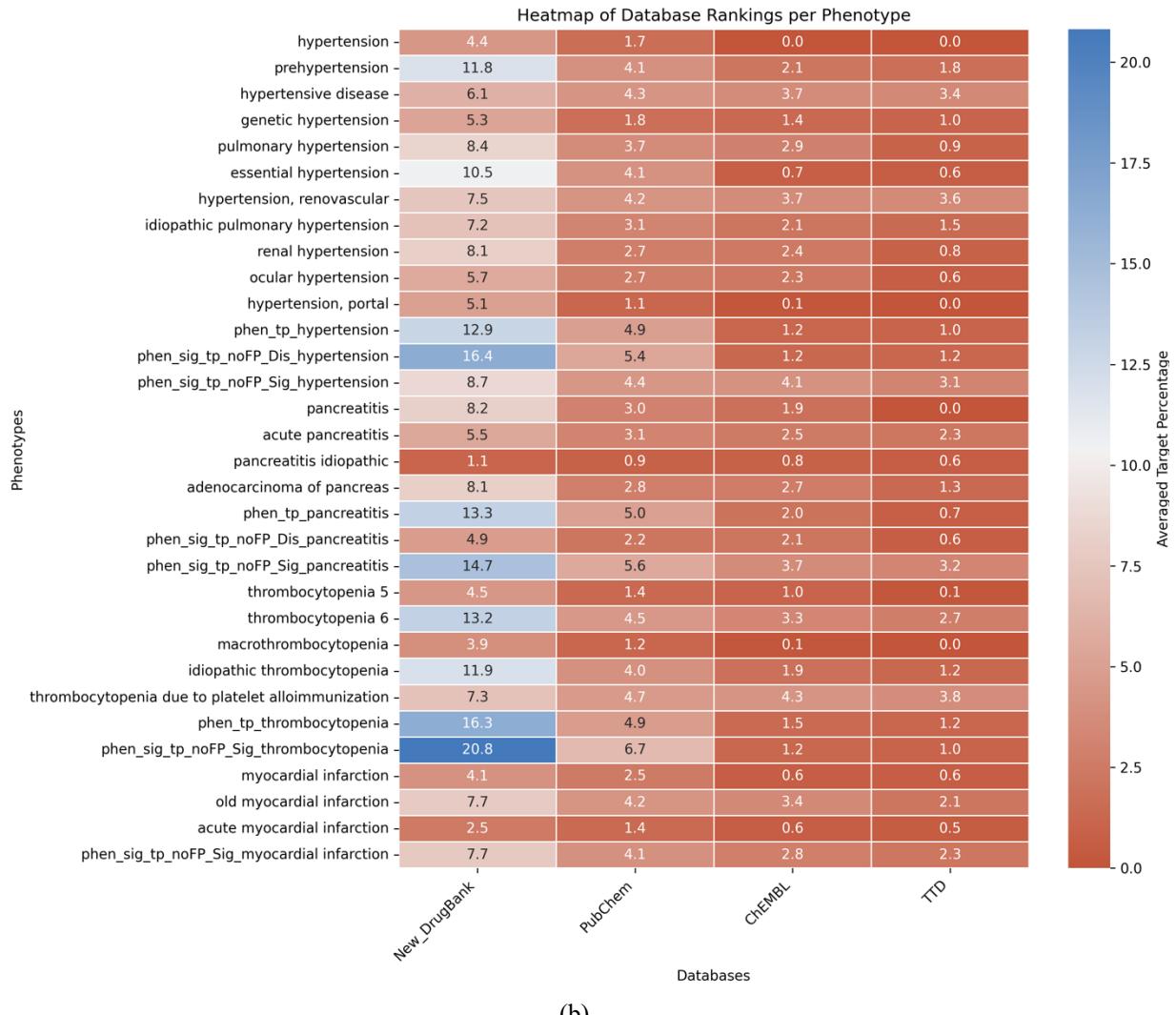


Figure S8. Heatmaps illustrating database rankings for phenotype prediction based on averaged target percentages in two filtered datasets. (a) Rankings for 111 drugs shared across TTD, New_DrugBank, ChEMBL, and STITCH (excluding PubChem and Pocket_Features). (b) Rankings for 147 drugs shared across TTD, PubChem, New_DrugBank, and ChEMBL (excluding STITCH and Pocket_Features). Blue shades indicate higher contributions of targets to phenotype prediction, while red shades signify lower contributions.