Analyzing Patterns of Computational Similarity between Kinase Ligands

Jack Ringer

July 2025

Abstract

Protein kinases are relevant to a large number of human pathologies, including cancer, immune disorders, and infectious diseases. An improved understanding of structure-activity relationships for kinase ligands would benefit the development of new treatments for these pathologies. The entire workflow developed as part of this project can be found on GitHub: https://github.com/Jack-42/ligandActivityAnalysis.

Introduction and Background

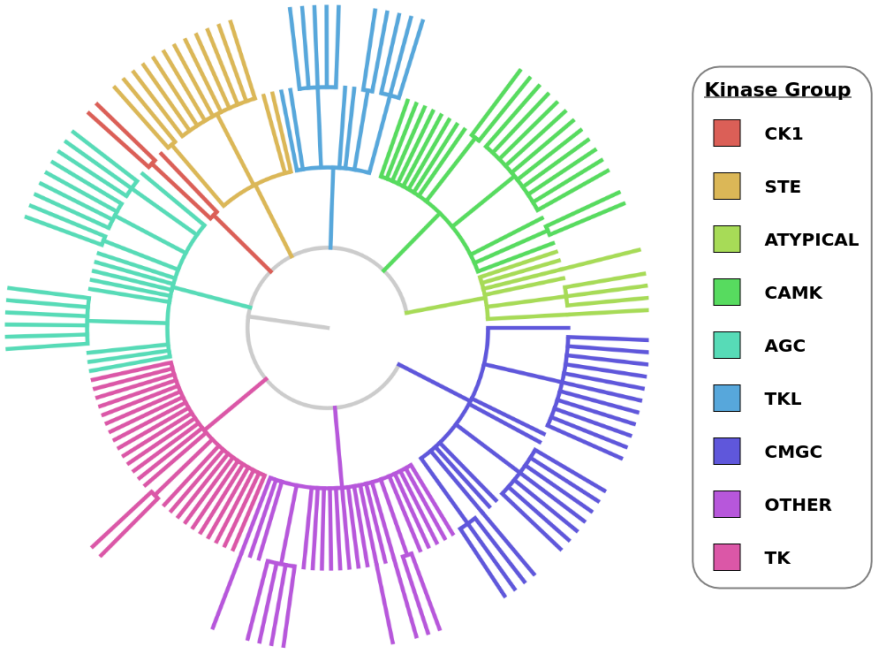

Figure 1: Phylogenetic Tree of the Human Kinome showing major groups as well as families and subfamilies. Generated using ETE4 with data from CheMBL.

Methodology

Results

Distribution

Figure 2 provides distributions for the $\binom{N}{2}$ similarity values per group (N = number of ligands), as well as the $\binom{9{,}995}{2} = 49{,}945{,}015$ similarity values calcualted for all kinase ligands in the dataset. Table 1

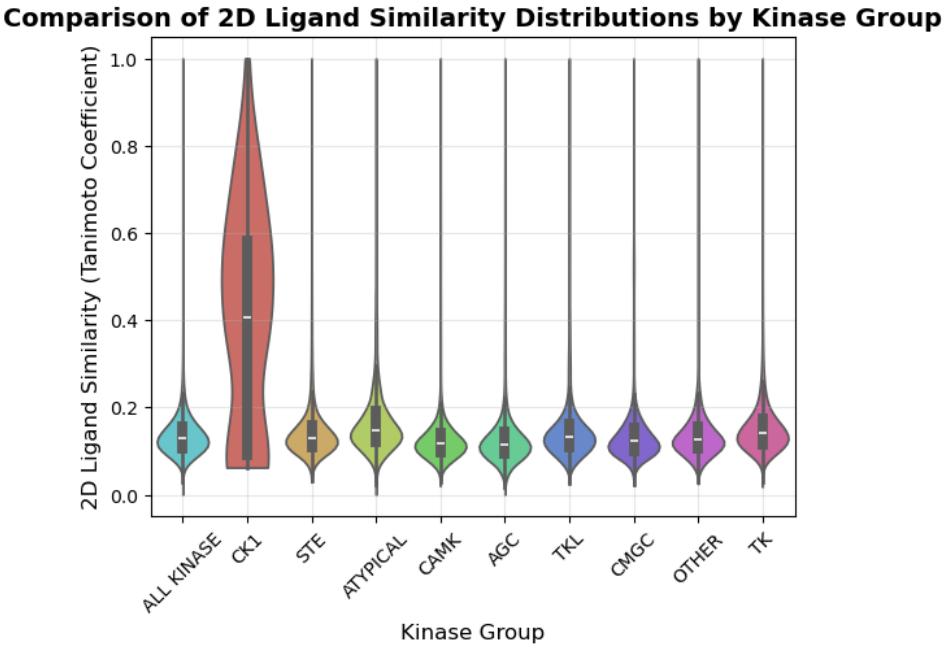

Figure 2: Comparison of 2D structural similarity distributions by kinase group

provides additional statistics for each group, as well as results from a Mann-Whitney U test (MWUT).

Kinase Group	N. Targets	N. Ligands	Group Median	Comparison Median	MWUT p-value
CK1	10	13	0.407	0.129	3.10×10^{-11}
STE	45	425	0.131	0.129	2.48×10^{-169}
ATYPICAL	15	357	0.149	0.129	$< 5 \times 10^{-324}$
CAMK	65	597	0.117	0.130	1.0
AGC	59	809	0.116	0.131	1.0
TKL	37	810	0.133	0.129	$< 5 \times 10^{-324}$
CMGC	58	1275	0.122	0.130	1.0
OTHER	56	727	0.127	0.129	1.0
TK	80	5347	0.140	0.121	$< 5 \times 10^{-324}$

Table 1: Table showing comparisons of targets, ligands, and ligand similarity distributions per kinase group. Shown p-values are calculated (with Bonferroni correction) from a MWUT where the alternative hypothesis is that the similarity values within the group are stochastically greater than the distribution of all similarity values.

Enrichment

Discussion

Conclusion

Acknowledgement

Much thanks to my advisor Dr. Vincent Metzger for his guidance and input over the course of this project. I'd also like to thank Dr. Jeremy Yang, Dr. Cristian Bologa, and Dr. Praveen Kumar for their feedback during weekly meetings over the course of the internship. Finally, I'd like to thank the authors of ChEMBL DB [1].

References

[1] Barbara Zdrazil, Eloy Félix, Fiona Hunter, Emma Manners, James Blackshaw, Sybilla Corbett, Marleen De Veij, Harris Ioannidis, David Méndez, Juan F Mosquera, María Paula Magariños, Nicolas Bosc, Ricardo Arcila, Tevfik Kizilören, Anna Gaulton, A. Patrícia Bento, Melissa F Adasme, Peter Monecke, Gregory A Landrum, and Andrew R Leach. The chembl database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Research*, 52(D1), Nov 2023.