Generative Topographic Mapping for Interpretable and Scalable DNA-Encoded Library Comparison

Louis Plyer[1], Alexey A. Orlov[1], Tagir N. Akhmetshin[1], Erik Yeghyan[1], Fanny Bonachera[1], Dragos Horvath[1] and Alexandre Varnek[1]

[1] Laboratory of Chemoinformatics, UMR 7140 CNRS, University of Strasbourg, 4, Blaise Pascal Str., 67000 Strasbourg, France

DNA-encoded libraries [1] (DELs) enable exploration of extremely large chemical spaces for drug discovery, yet efficient comparison and prioritization of DEL collections remain challenging due to their scale. In this work [2], we investigate Generative Topographic Mapping [3] (GTM) as a scalable and interpretable framework for DEL similarity analysis. DELs were represented as GTM responsibility vectors, enabling rapid comparison of entire libraries without exhaustive compound pairwise similarity calculations. Using 100 diverse DEL subsets and a reference set of cyclin-dependent kinase 2 (CDK2) ligands from ChEMBL, we benchmarked GTM-based metrics against Morgan count fingerprint-based similarity approaches. The proposed metrics showed strong agreement with descriptor-space methods, achieving Spearman rank correlations of 0.7 and enrichment factors at 5% (EF5%) up to 12 for retrieval of top-ranked DELs. In addition, GTM generated intuitive 2D landscapes that facilitate visualization of library overlap, diversity, and target-oriented enrichment. These results demonstrate that GTM provides an efficient and explainable framework for large-scale DEL profiling and prioritization.

Bibliography:

[1] A. Gironda-Martínez, E. J. Donckele, F. Samain, and D. Neri, “DNA-Encoded Chemical Libraries: A Comprehensive Review With Succesful Stories and Future Challenges,” ACS Pharmacology & Translational Science 4, no. 4 (2021): 1265–1279.
[2] Plyer, L. et al. Interpretable and Scalable Similarity Metrics for DNA-Encoded Library Design Using Generative Topographic Mapping. Molecular Informatics 45, e70026 (2026).
[3] N. Kireeva, I. I. Baskin, H. A. Gaspar, D. Horvath, G. Marcou, and A. Varnek, “Generative Topographic Mapping (GTM): Universal Tool for Data Visualization, Structure-Activity Modeling and Dataset Comparison,” Molecular Informatics 31, no. 3–4 (2012), 301–312.