Chemical Space Visualization Toward Data-Driven DNA-Encoded Library Design

Kazuma Kaitoh, Nagoya University

DNA-encoded library (DEL) is a new technology for hit compound identification. Each compound in the DEL has a specific DNA-tag for identifying the chemical structure. The DEL is much larger than traditional chemical libraries. Thus, the DEL is suitable for performing large-scale compound screening. However, because of the huge chemical space, a rational DEL design is crucial for efficient identification of hit compounds. The chemical space visualization is a powerful tool for designing the DEL and obtaining the chemical features. In this study, we proposed the imaginary DEL (iDEL) for the chemical space visualization of compounds in the DEL. The iDEL was composed of chemical structures based on the combinations of scaffolds and fragments. The chemical structures were represented as fingerprints, physicochemical properties, and the latent vectors of Transformer. The dimensionality reduction was executed by PCA and UMAP. We will show the results of the iDEL-based visualization and discuss their features on the chemical space.