Representation learning and graph neural networks for polyphamacology: application to combinatorial code of olfaction

Matej Hladis, Maxence Lalis, Sebastien Fiorucci, Jeremie Topin

Institut de Chimie de Nice (ICN), Université Côte D'Azur, CNRS, France

Keywords: protein-molecule interaction, olfaction, graph neural networks, protein language modelling, GPCR, olfactory receptor

Mammalian sense of smell can distinguish a myriad of various odors using a combinatorial coding scheme, in which different odors are represented by the activity patterns of hundreds of proteins, called olfactory receptors (ORs). Each odorant molecule activates a set of these ORs, creating a representation that our brain eventually interprets as a perception, which we call smell. However, revealing this combinatorial code is a long-standing challenge and determining the code even for a single molecule, is costly and time-consuming. For humans, nearly 400 laboratory experiments are required for each molecule. In this work, we combine protein language with graph neural networks to predict OR activation, and propose a tailored architecture incorporating inductive biases from the protein-molecule interaction. On a novel dataset of 46 700 OR-molecule pairs, this model outperforms state-of-the-art drug-target interaction prediction models as well as standard GNN baselines. Notably, our predictions are in agreement with combinatorial coding theory in olfaction. Our results reveal consistent coding for a large number of odor families and the model suggests new insights such as previously unknown pairs of enantiomers with distinct combinatorial codes.