

# FIDDLE: An integrative deep learning framework for functional genomic data inference

## Abstract

Numerous advances in sequencing technologies have revolutionized genomics through generating many types of genomic functional data. Statistical tools have been developed to analyze individual data types, but there lack strategies to integrate disparate datasets under a unified framework. Moreover, most analysis techniques heavily rely on feature selection and data preprocessing which increase the difficulty of addressing biological questions through the integration of multiple datasets. Here, we introduce FIDDLE (Flexible Integration of Data with Deep LEarning) an open source data-agnostic flexible integrative framework that learns a unified representation from multiple data types to infer another data type. As a case study, we use multiple *Saccharomyces cerevisiae* genomic datasets to predict global transcription start sites (TSS) through the simulation of TSS-seq data. We demonstrate that a type of data can be inferred from other sources of data types without manually specifying the relevant features and preprocessing. We show that models built from multiple genome-wide datasets perform profoundly better than models built from individual datasets. Thus FIDDLE learns the complex synergistic relationship within individual datasets and, importantly, across datasets.