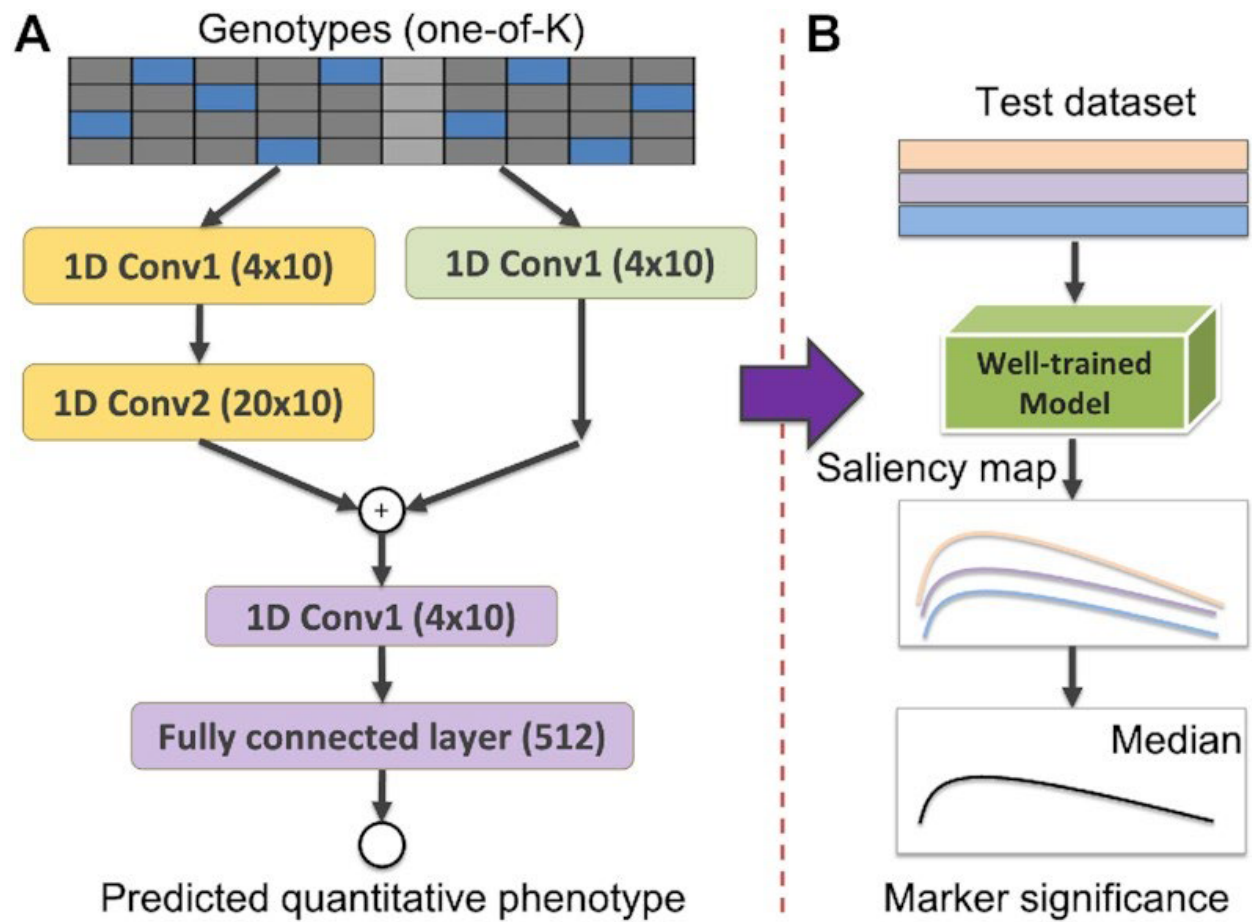


Shuai Zeng, Ziting Mao, Yijie Ren, Duolin Wang, Dong Xu, Trupti Joshi. G2PDeep: a web-based deep-learning framework for quantitative phenotype prediction and discovery of genomic markers. *Nucleic Acids Research*, 49:W228-W236, 2021.

G2PDeep is an open-access web server, which provides a deep-learning framework for quantitative phenotype prediction and discovery of genomics markers. It uses zygosity or single nucleotide polymorphism (SNP) information from plants and animals as the input to predict quantitative phenotype of interest and genomic markers associated with phenotype. It provides a one-stop-shop platform for researchers to create deep-learning models through an interactive web interface and train these models with uploaded data, using high-performance computing resources plugged at the backend. G2PDeep also provides a series of informative interfaces to monitor the training process and compare the performance among the trained models. The trained models can then be deployed automatically. The quantitative phenotype and genomic markers are predicted using a user-selected trained model and the results are visualized. Our state-of-the-art model has been benchmarked and demonstrated competitive performance in quantitative phenotype predictions by other researchers. In addition, the server integrates the soybean nested association mapping (SoyNAM) dataset with five phenotypes, including grain yield, height, moisture, oil, and protein. A publicly available dataset for seed protein and oil content has also been integrated into the server. The G2PDeep server is publicly available at <http://g2pdeep.org>. The Python-based deep-learning model is available at https://github.com/shuaizengMU/G2PDeep_model.



(A) Architecture of dual-stream CNN model. The genotypes are one-hot coded. (B) Flowchart of genomic markers discovery using a well-trained model and saliency map.