

Brain-ResNet

This manuscript ([permalink](#)) was automatically generated from [sq-96/Brain-ResNet@f9202b5](#) on April 12, 2020.

Authors

- Sheng Qian

Abstract

Decoding the regulatory behavior of DNA sequences and the functional effects of noncoding variants is a preeminent challenge in understanding the mechanisms of gene regulation. This is also important for the genetics of common diseases, as most disease-associated variants are located in noncoding regions of the genome. Recently, Convolutional Neural Networks (CNNs) based methods have been developed to predict genome-wide chromatin profiles in various cellular contexts. However, these tools and resources were often trained in cell lines or bulk tissues that are not necessarily disease-related. This is particularly an issue for neuropsychiatric disorders, where the most relevant cell and tissue types are missing in the training data used by current tools. Here we introduce a Convolutional Residual Networks (ResNet) based architecture that trains with user-provided epigenomic profiles in specific cell types while borrowing information from an extensive compendium of publicly available chromatin profiles via transfer learning. We applied this framework to a collection of epigenomic data representing early neurodevelopment, including data from iPS cell derived neurons and fetal brain. We show that our method predicts chromatin accessibility of Induced Pluripotent Stem Cell (iPSC) derived neuronal cells with higher accuracy than previous methods. Using a list of experimentally derived chromatin accessibility variants, we show that our models can predict regulatory variants with better accuracy than publicly available noncoding variant scoring tools. We also found that with the transfer learning framework, we can recover cell-type-specific motifs from our architecture, shedding light on possible regulatory mechanisms in specific cell types. In summary, we provide a general deep transfer learning framework to train on user-provided epigenomic data, and our resources would be useful for studying neuropsychiatric diseases.

Introduction

Next-generation sequencing(NGS) technologies have given rise to the development of many sequencing assays such as ATAC-seq[7], DNase-seq[49], ChIPseq[39], RNA-seq[58], and FIAR-seq[19] that measure the epigenomic landscapes across many cellular contexts, including histone marks, TF binding and chromatin accessibility. These epigenomic annotations aid the characterization of noncoding genomic variants and show promises in assessing disease-associated variants and understanding the underlying transcription machinery[31, 53]. There has been a joint effort to survey the noncoding part of the human genome by the community, and numerous noncoding genomic sites have been statistically identified for association with complex traits[11]. Leveraging these resources, researchers have developed machine learning models to learn features of DNA sequences that predict chromatin profiles such as protein binding sites, chromatin accessibility, histone marks and methylation of DNA sequences[2, 30, 4, 40, 13, 43, 61]. Once a sequence based model is trained to predict a certain epigenomic feature, a researcher can use it to predict the likely epigenomic effect of a DNA variant.

Results

Citations

Citation by DOI [[1](#)].

Figures

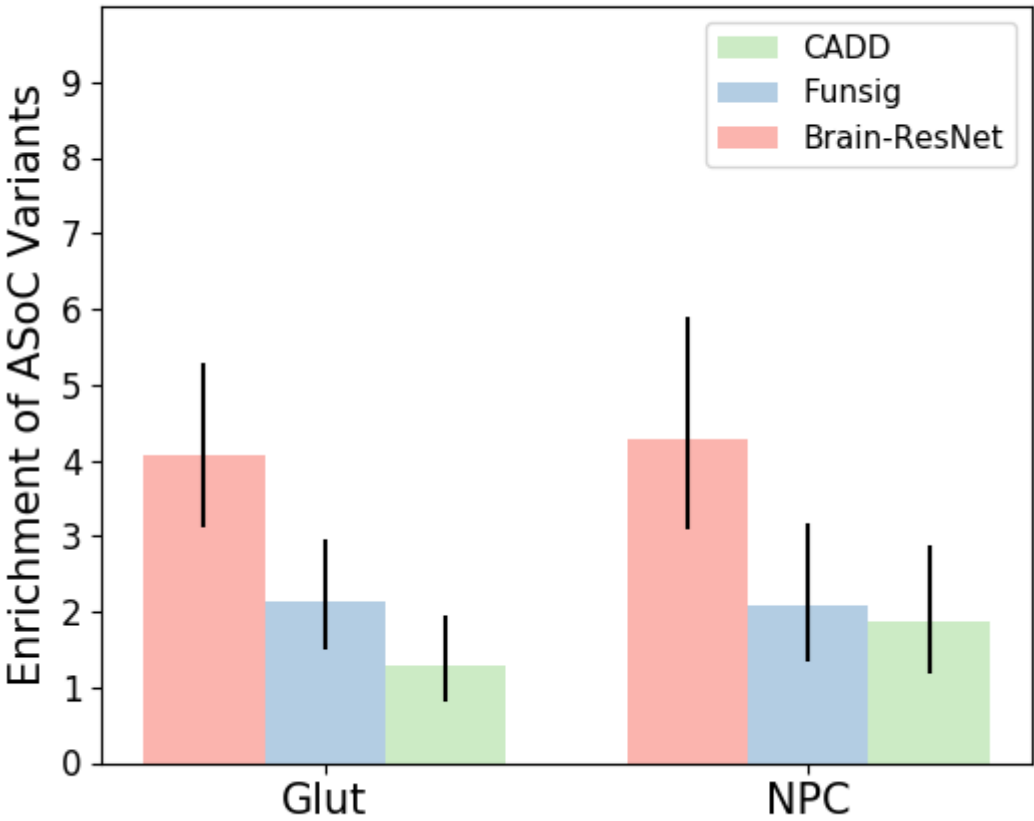


Figure 1: ASoC Enrichment. Loaded from the latest version of image on GitHub.

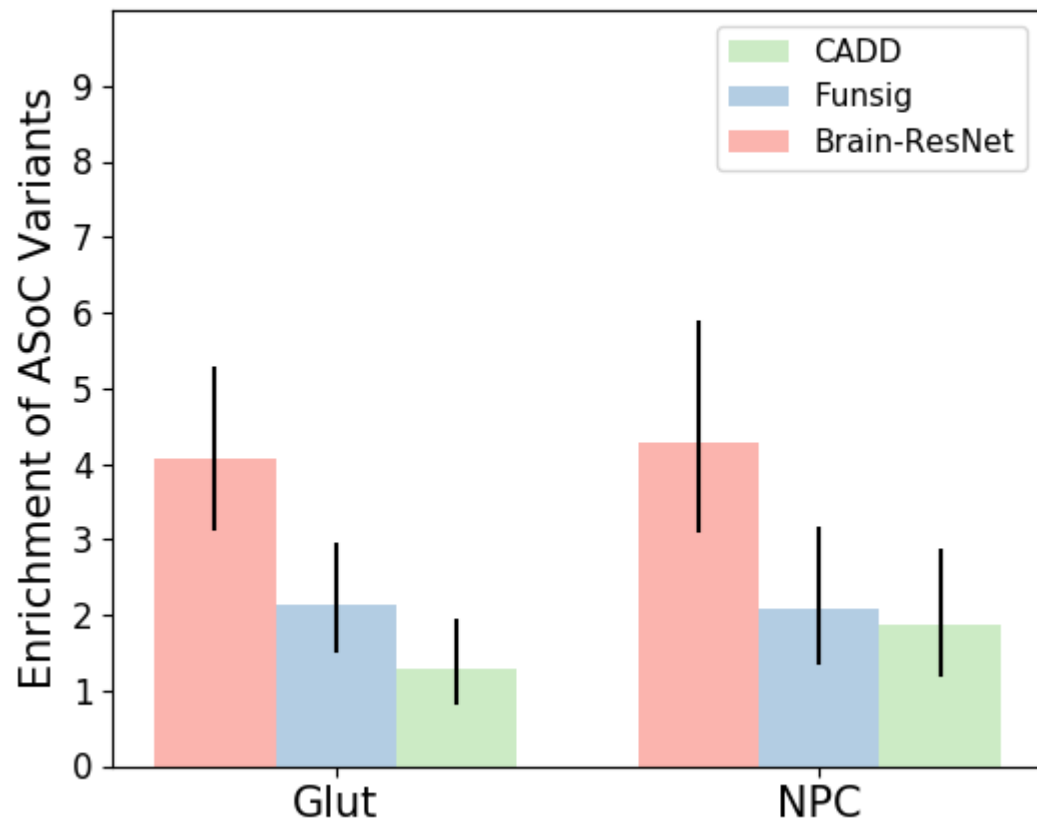


Figure 2: GERP Score Distribution. Loaded from the latest version of image on GitHub.

References

1. Leveraging genetic interactions for adverse drug-drug interaction prediction

Sheng Qian, Siqi Liang, Haiyuan Yu

PLOS Computational Biology (2019-05-24) <https://doi.org/ggrj6t>

DOI: [10.1371/journal.pcbi.1007068](https://doi.org/10.1371/journal.pcbi.1007068) · PMID: [31125330](https://pubmed.ncbi.nlm.nih.gov/31125330/) · PMCID: [PMC6553795](https://pubmed.ncbi.nlm.nih.gov/PMC6553795/)