SCIENCE MEETS LIFE

Deep learning of Drosophila brain enhancers leveraged by single-cell RNA and ATAC-seq

Ibrahim I. Taskiran, Jasper Janssens, Sara Aibar, Stein Aerts

Ibrahim I. Taskiran

PhD Student

Laboratory of Computational Biology (Stein Aerts Lab)

VIB - KU Leuven - Belgium

VIB

KU LEUVEN

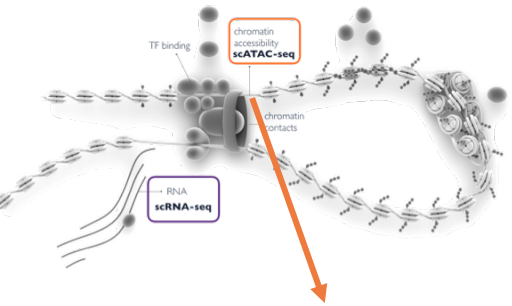
VIB Conference

Emerging Technologies in Single Cell Research

19-20 NOVEMBER 2020

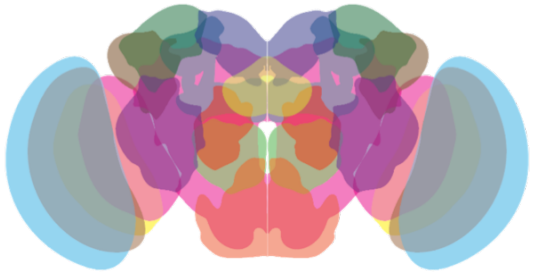
Overview

Enhancer Code



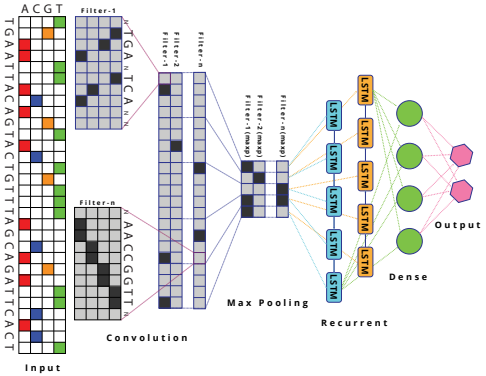
Single-cell ATAC-seq and single-cell RNA-seq provide new opportunities to study gene regulation in heterogeneous cell populations such as complex tissues or dynamic processes. In these complex cell populations, cell identity is defined by the activity of genes that are regulated by transcription factors (TFs) binding to cis-regulatory elements on the DNA

Drosophila Brain

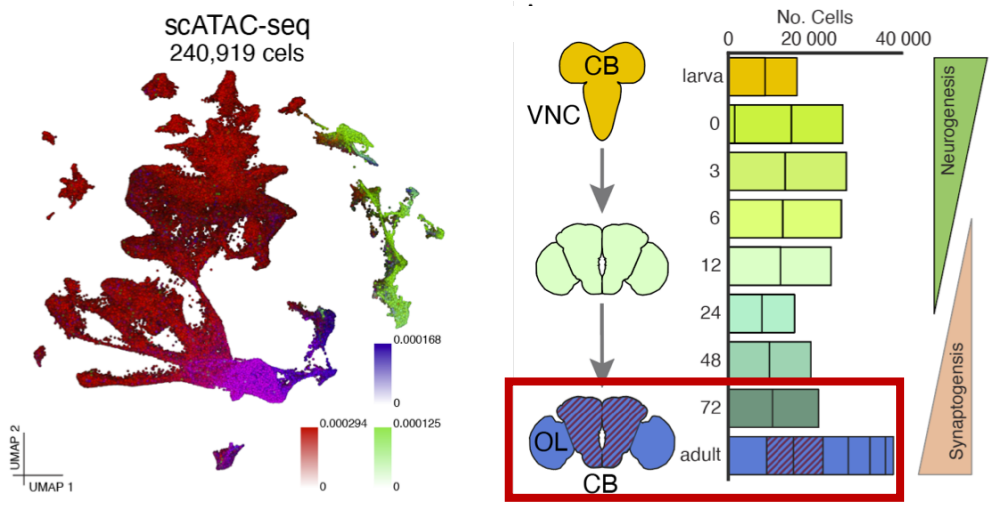


We apply deep learning approaches to analyze sets of co-accessible enhancers identified through scATAC-seq data by topic modeling, with the goal to predict the spatiotemporal pattern of enhancer accessibility directly from the enhancer sequence. We trained Convolutional-Recurrent Neural Networks on several neuronal cell types and validated our predictions

Deep Learning

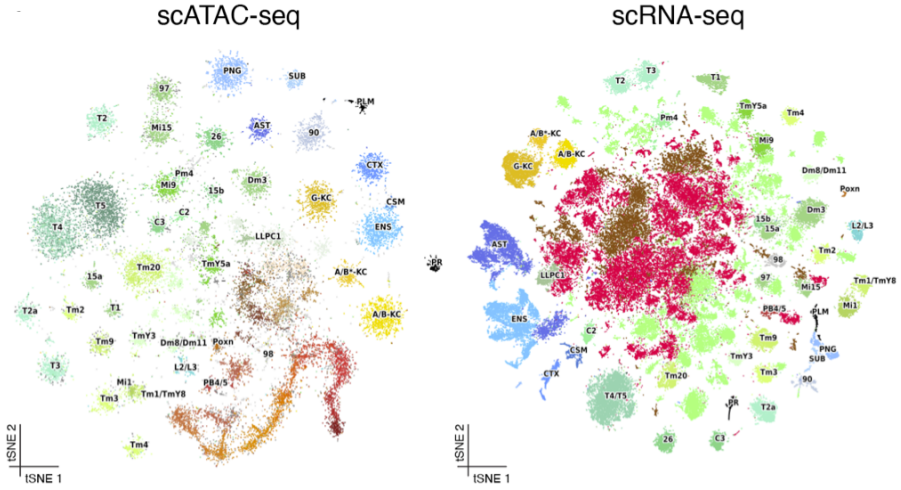


Single-cell Atlas of the Developing Fly Brain



scATAC-seq is performed on developing fly brain

We analyzed around 240,000 cells that we profiled with scATAC-seq on developing fly brain and we identified 79 distinct clusters of cell types on adult brain



scATAC-seq and scRNA-seq is co-clustered on adult fly brain

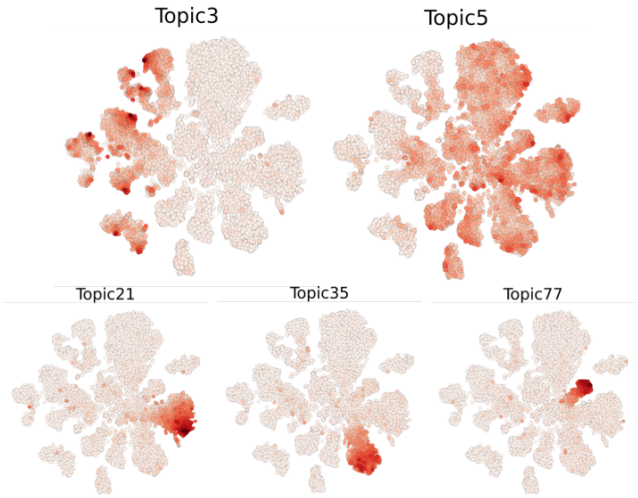
By using the scRNA-seq profiling of the Drosophila brain that we analyzed in previous studies (Davie et al., 2018), we integrated scATAC and scRNA data where 43 of 79 distinct clusters of cell types are linked cell clusters in the scRNA-seq data

Deep Learning on Fly Brain

1 - Glia, KC, and T neurons are further clustered

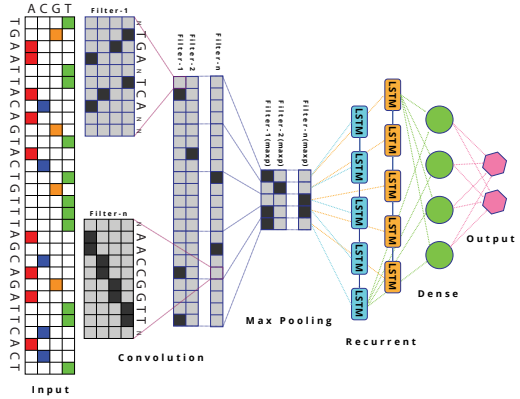
scATAC-seq profile of 5 Glia, 5 Kenyon Cell (KC), and 5 T-Neuron sub-types are selected and they are further clustered by cisTopic (Bravo et al., 2019)

2 - Cell-type specific topics (DARs) are identified



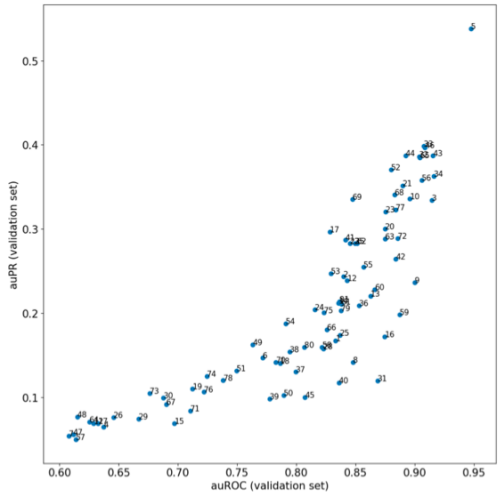
We identified cell-type specific accessible regions. Here you see specifically accessible regions for glia cells, neurons, and 3 KC subtypes

3 - Deep Learning model is trained



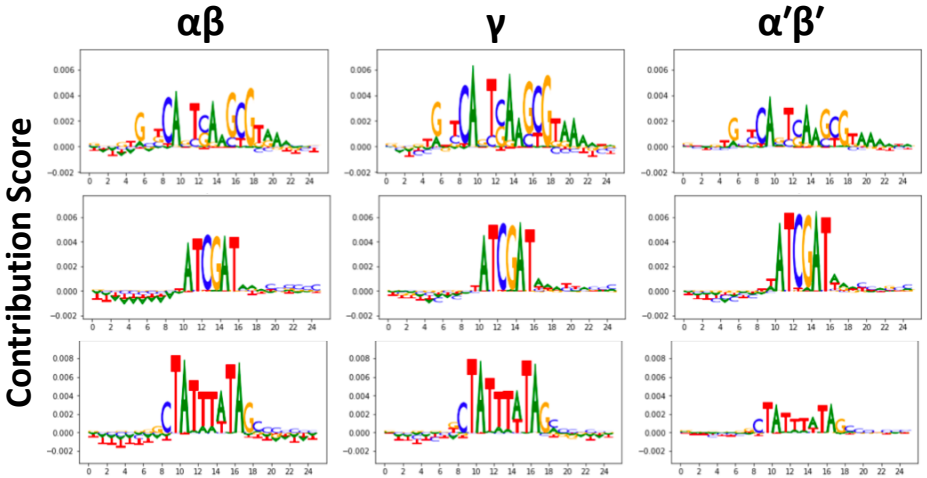
Deep learning model is trained on the output of cisTopic clustering to predict cell-type specific accessibility from DNA sequence

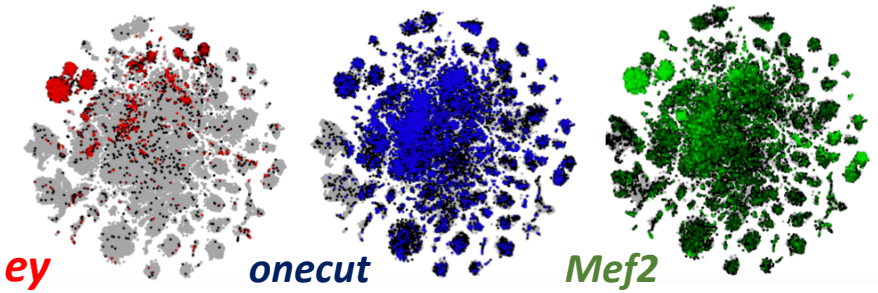
4 - Performance of the model



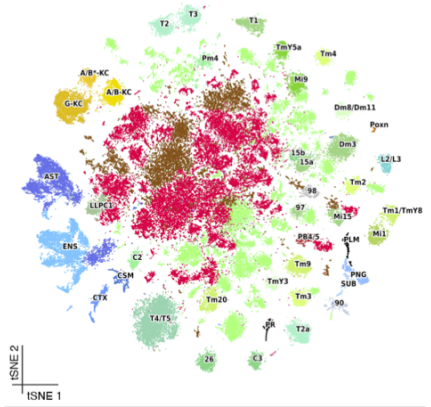
Model performed very well on many of the topics. From this point we started to analyze the model to understand the enhancer code

5 – Key transcription factors for the KCs are identified



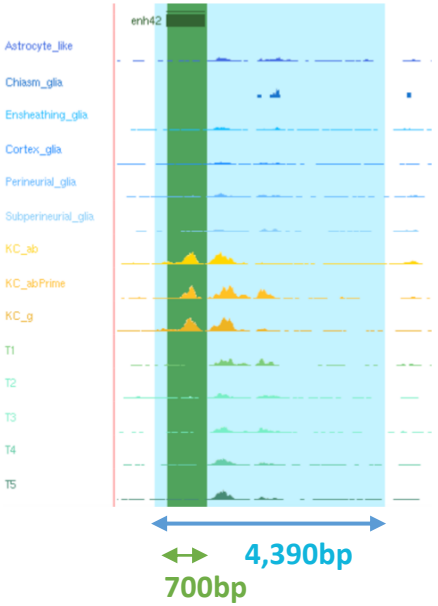


scRNA-seq

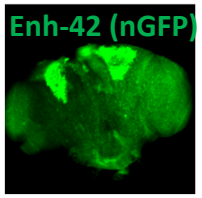


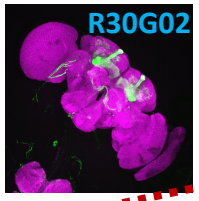
By assigning filters to KC specific topics we identified key transcription factors (TFs) from filters and also by using TF-MoDISco (Shrikumar et al., 2018). Identified TFs matched very well with the TF expression profiles in the single-cell RNA-seq data.

Combinatorial Code of Kenyon Cell Enhancers

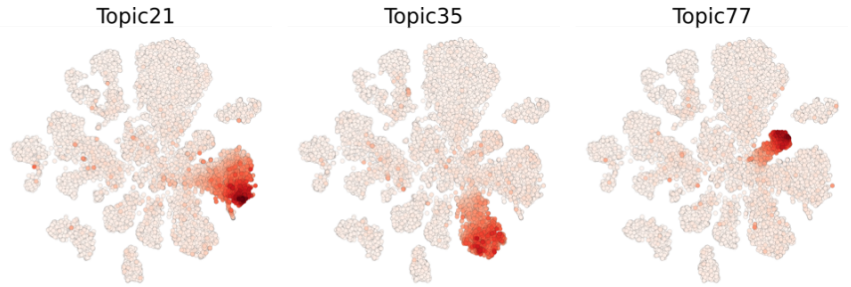


Here you see an example region of 4,390bp (blue), which was tested with FlyLight project. It gives a GFP expression on KCs. With the help of scATAC-seq signal and the deep learning model we identified a region shorter than 700bp (green), which is enough to drive GFP expression on KCs. The model prediction score on that regions says that it is a KC specific accessible region. Then, when we looked at the DeepExplainer (Lundberg et al., 2017) profile, we identified combinatorial binding of KC specific key transcription factors on a single enhancer. Further that, in silico saturation mutagenesis assay shows the effect of single nucleotide change on the enhancer.

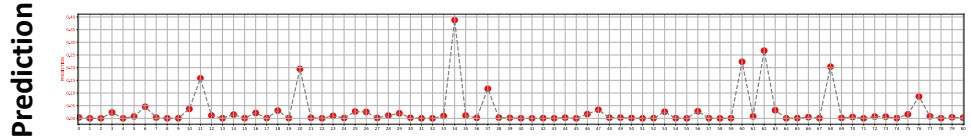


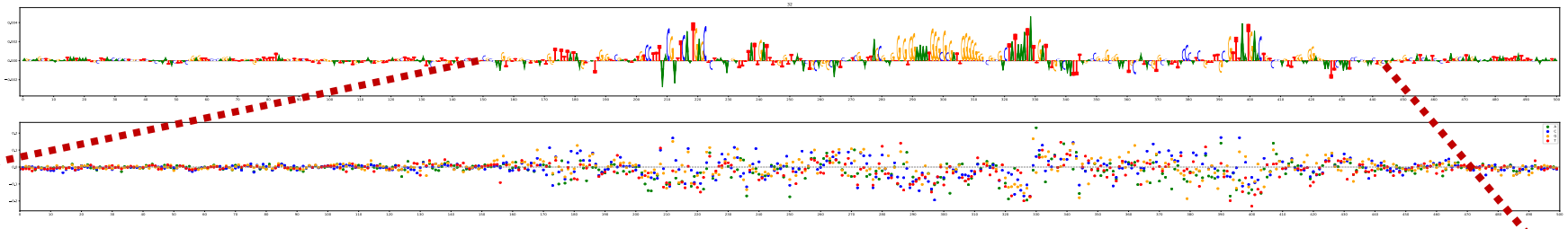


FlyLight



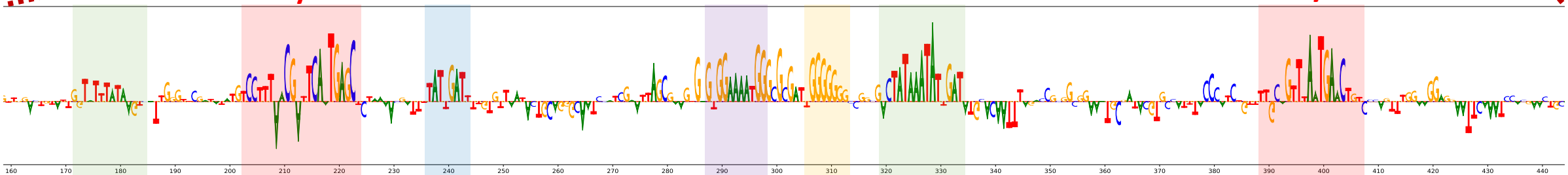
Prediction





Mef2 ey onecut NFAT sr Mef2 ey

Deep Explainer



In silico saturation mutagenesis