[BIOINFO 2024]

scDeepLUCIA: a deep-learning model to elucidate cell-type specific 3D gene regulation from low-resolution single-cell 3D genome information

Dongchan Yang[1], and Inkyung Jung[1,*]

[1]*Department of Biological Sciences, Korea Advanced Institute of Science and Technology (KAIST)*
**Corresponding author: ijung@kaist.ac.kr*

Recently developed single-cell HiCAR (scHiCAR) joint profiles transcriptome, chromatin accessibility, and 3D genome of the individual cell. However, comparative 3D-multiomic analysis is challenging as the number of chromatin loops called by conventional chromatin loop callers strongly depends on the library depth. scDeepLUCIA is developed to address this issue by combining single-cell 3D-multiomics information using a deep neural network. scDeepLUCIA consists of 4 modules: the genome sequence motif module, the chromatin accessibility module, the low-depth contact matrix module, and the integration module. The genome sequence motif module learns the genomic sequence motifs which are responsible for the chromatin loop formation. The chromatin accessibility module learns the chromatin accessibility to discriminate the chromatin loop anchors. The low-depth contact matrix module augments the low-depth contact matrix with 2D convolution. Finally, the integration module integrates the sequence motif, chromatin accessibility, and contact matrix to determine the chromatin loop. scDeepLUCIA is trained with a simulated scHiCAR dataset. After training, scDeepLUCIA can accurately identify the chromatin loops with AUROC > 0.93 even though the trainset and test set are from different species and different cell types. Moreover, the chromatin loops called by scDeepLUCIA are enriched with eQTL SNP and chromatin accessibility peaks. Then, scDeepLUCIA is applied to predict chromatin loops from the mouse brain scHiCAR dataset. Even greater than 50 50-fold change between the number of cells from the hugest (L23 IT 1) and smallest (L5 ET), scDeepLUCIA identified a comparable number (around 40,000 loops) of chromatin loops for each mouse brain cell cluster. Then, the validation of the chromatin loops with the VISTA enhancer, and gene expression revealed that the pairs of distal *cRE*s and genes with scDeepLUCIA loops are enriched with VISTA enhancer and highly expressed respectively. Interestingly, the correlation between gene expression and distal *cRE* activity is high only if the pairs of *cRE*s and genes are mediated by the chromatin loops of exact cell types. Moreover, the gene ontology (GO) analysis of the target gene reflects the cell identity.