[BIOINFO 2023]

Development of large-scale structural variation detection software in Hi-C and single-cell Hi-C using few-shot learning

Kyukwang Kim and Inkyung Jung*

Department of Biological Sciences, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea

**Corresponding author: ijung@kaist.ac.kr*

Disruption of 3D chromatin structure due to large-scale structural variation can be identified by using Hi-C technique. The utilization of the Hi-C not only enables the detection of the SV event, but also allows the interpretation of the SV's impact in terms of rearrangement of regulatory elements. To understand the cancer genome in terms of 3D genome, several Hi-C based SV detection software have been developed. As Hi-C contact maps contains signals from various sources, software that uses deep learning was also developed to successfully find SV patterns in the Hi-C contact maps. Variants of Hi-C techniques such as a single-cell Hi-C have been developed to obtain additional information other than the merged 3D chromatin structure. However, the conventional deep learning has poor classification ability for untrained data and requires a lot of data for re-training, which hinders SV detection from the new experimental technique data produced in small quantities. To solve this problem, we applied a few-shot learning method, which enables a pre-trained model to be trained using only a few labeled samples per class. The developed software was trained using Hi-C data from 9 cancer cell lines and 4 normal cell lines/tissues. In the benchmark using colorectal cancer patient tumor/normal tissue Hi-C data showed a comparable true positive rate (1.03 fold higher) with the existing SOTA software (EagleC), while 2.71 fold less false positive calls were made. A benchmark on single-cell Hi-C showed an accuracy of 87.5% after the fine-tuning was conducted. The development of our method expands the SV detection to single cell Hi-C data, which deepens our observation of the cancer 3D genome.