

Evaluasi Metode Timesvector Triclustering Menggunakan Gene Ontology pada Data Ekspresi Gen Human Embryonic Stem Cell = Evaluation of TimesVector Triclustering Method Using Gene Ontology on Human Embryonic Stem Cell Gene Expression Data

Latif Raditya Rusdi, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920563986&lokasi=lokal>

Abstrak

Triclustering merupakan salah satu teknik data mining yang bertujuan untuk mengelompokkan data berbentuk tiga dimensi secara simultan. Salah satu pendekatan yang digunakan dalam triclustering adalah pendekatan pattern-based, contohnya Timesvector. Metode timesvector dirancang khusus untuk pengelompokan data deret waktu tiga dimensi yang bertujuan menangkap pola ekspresi gen yang sama atau berbeda antara dua atau lebih kondisi eksperimen. Implementasi metode timesvector dilakukan pada data ekspresi gen human embryonic stem cell (H1-hESC) yang diberi protein morfogenetik tulang (BMP4) dan dikondisikan di dalam ruang dengan tingkat oksigen 4% dan 20, serta diamati pada 6 titik waktu berbeda selama 120 jam. Triclustering dilakukan dengan lima skenario menggunakan cluster sejumlah 257 dan threshold yang berbeda. Berdasarkan skenario tersebut, metode timesvector menghasilkan skenario terbaik pada skenario dengan threshold 1,5 yang menggunakan validasi berdasarkan nilai coverage. Berdasarkan hasil skenario terbaik, dihasilkan 9 pola DEP, 24 pola ODEP, dan 37 pola SEP dan dari pola tersebut dilakukan analisis Gene Ontology (GO) untuk mengukur kualitas tricluster dalam penggambaran konsep GO. Analisis GO menggunakan Database for Annotation, Visualization, and Integrated Discovery (DAVID) tools untuk menghitung nilai p-value. Pada analisis GO dipilih p-value terkecil pada pola DEP, ODEP, dan SEP sebagai tricluster terbaik, yaitu DEP pada tricluster ke 8, ODEP pada tricluster ke-1, dan SEP pada tricluster ke-26. Berdasarkan tricluster terbaik pada pola DEP dan ODEP dapat dikatakan bahwa kondisi oksigen tingkat fisiologis 4 % dan tingkat atmosfer 20 % memiliki perbedaan dalam mengidentifikasi gen kandidat pada H1-hESC yang mampu berdiferensiasi menjadi trofoblas, sedangkan SEP tidak memiliki perbedaan dalam mengidentifikasi gen kandidat pada H1-hESC dengan dua kondisi berbeda.

.....Triclustering is one of the data mining techniques that aims to cluster three-dimensional data simultaneously. One of the approaches used in triclustering is a pattern-based approach, such as Timesvector. The timesvector method is specifically designed for clustering three-dimensional time series data that aims to capture gene expression patterns that are the same or different between two or more experimental conditions. The implementation of the timesvector method was performed on human embryonic stem cell (H1-hESC) gene expression data treated with bone morphogenetic protein (BMP4) and conditioned in a chamber with 4% and 20 oxygen levels and observed at 6 different time points for 120 hours. Triclustering was performed with five scenarios using 257 clusters and different thresholds. Based on these scenarios, the timesvector method produces the best scenario in the scenario with a threshold of 1.5 which uses validation based on the coverage value. Based on the results of the best scenario, 9 DEP patterns, 24 ODEP patterns, and 37 SEP patterns were generated from these patterns. Gene Ontology (GO) analysis was carried out to measure the quality of the tricluster in describing the GO concept. GO analysis uses Database for Annotation, Visualization, and Integrated Discovery (DAVID) tools to calculate the p-value. In the GO analysis, the smallest p value in the DEP, ODEP, and SEP patterns was selected as the best

tricluster, namely DEP in the 8th tricluster, ODEP in the 1st tricluster, and SEP in the 26th tricluster. Based on the best tricluster in the DEP and ODEP patterns, it can be said that the oxygen conditions of 4% physiological level and 20% atmospheric level have differences in identifying candidate genes in H1-hESC that are able to differentiate into trophoblasts, while SEP has no difference in identifying candidate genes in H1-hESC with two different conditions.