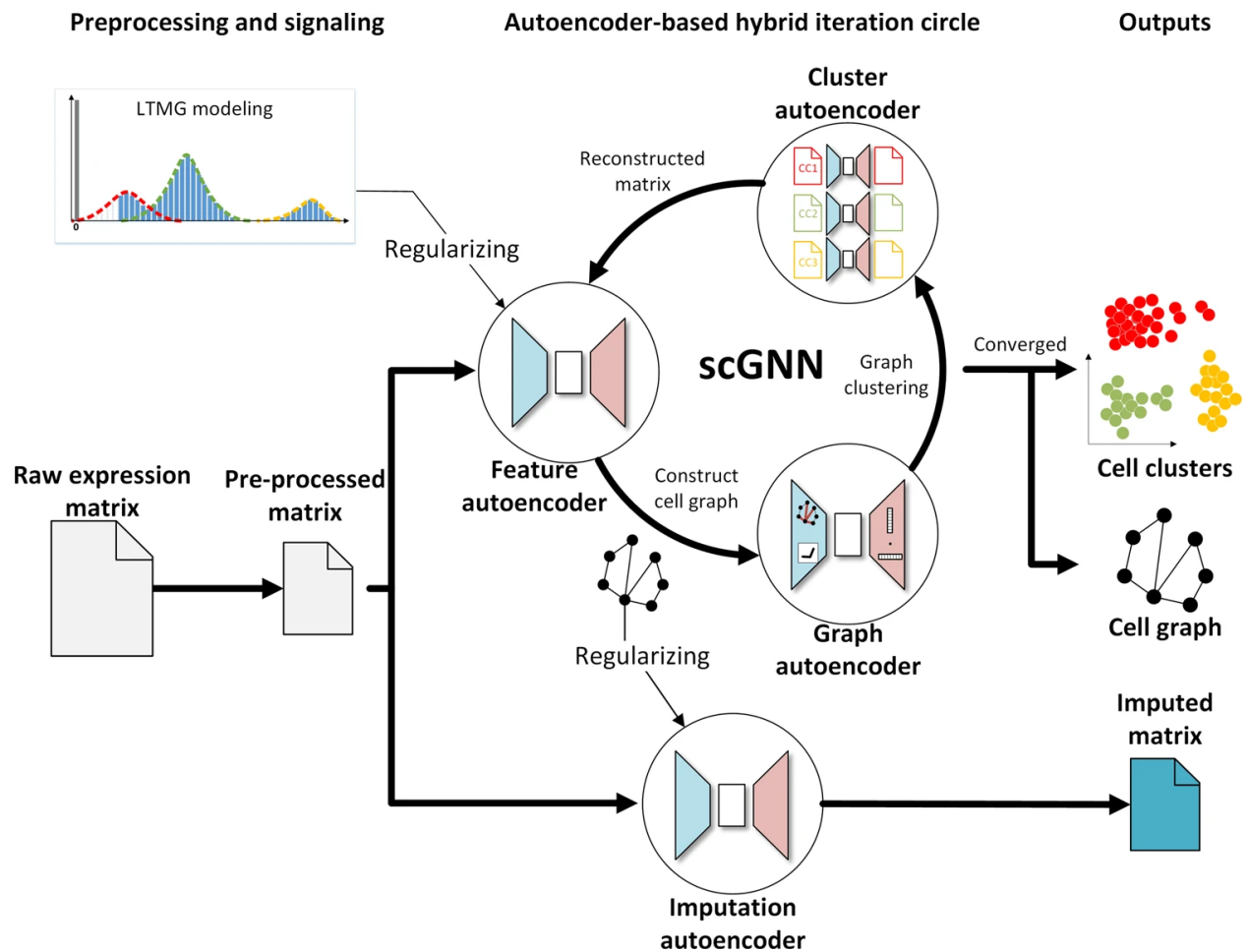


Juexin Wang, Anjun Ma, Yuzhou Chang, Jianting Gong, Yuexu Jiang, Ren Qi, Cankun Wang, Hongjun Fu, Qin Ma, Dong Xu. scGNN is a novel graph neural network framework for single-cell RNA-Seq analyses. *Nature Communications*. 12:1882, 2021.

Single-cell RNA-sequencing (scRNA-Seq) is widely used to reveal the heterogeneity and dynamics of tissues, organisms, and complex diseases, but its analyses still suffer from multiple grand challenges, including the sequencing sparsity and complex differential patterns in gene expression. We introduce the scGNN (single-cell graph neural network) to provide a hypothesis-free deep learning framework for scRNA-Seq analyses. This framework formulates and aggregates cell–cell relationships with graph neural networks and models heterogeneous gene expression patterns using a left-truncated mixture Gaussian model. scGNN integrates three iterative multi-modal autoencoders and outperforms existing tools for gene imputation and cell clustering on four benchmark scRNA-Seq datasets. In an Alzheimer’s disease study with 13,214 single nuclei from postmortem brain tissues, scGNN successfully illustrated disease-related neural development and the differential mechanism. scGNN provides an effective representation of gene expression and cell–cell relationships. It is also a powerful framework that can be applied to general scRNA-Seq analyses.



The architecture of scGNN. It takes the gene expression matrix generated from scRNA-Seq as the input. LTMG can translate the input gene expression data into a discretized regulatory signal as the regularizer for the feature autoencoder. The feature autoencoder learns a dimensional representation of the input as embedding, upon which a cell graph is constructed and pruned. The graph autoencoder learns a topological graph embedding of the cell graph, which is used for cell-type clustering. The cells in each cell type have an individual cluster autoencoder to reconstruct gene expression values. The framework treats the reconstructed expression as a new input iteratively until converging. Finally, the imputed gene expression values are obtained by the feature autoencoder regularized by the cell-cell relationships in the learned cell graph on the original pre-processed raw expression matrix through the imputation autoencoder. LTMG is abbreviated for the left-truncated mixed Gaussian model.