Supplementary Information for “Persistent spectral simplicial complex-based machine learning (PerSpectSC-ML) for chromosomal structural analysis in cellular differentiation”

Weikang Gong, JunJie Wee, Min-Chun Wu, Xiaohan Sun, Chunhua Li and Kelin Xia

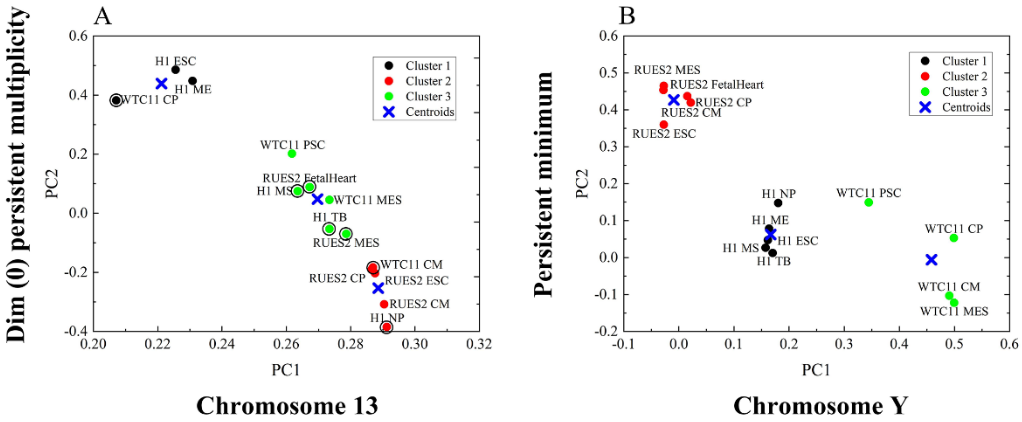

Figure S1. Performance comparison of PerSpectSC-ML's (here, ML method is PCA-assisted *k*-means) results obtained on Dim (0) persistent multiplicity of chromosome 13 (**A**) and persistent minimum of chromosome Y (**B**). Miscla-ssified cells are circled.

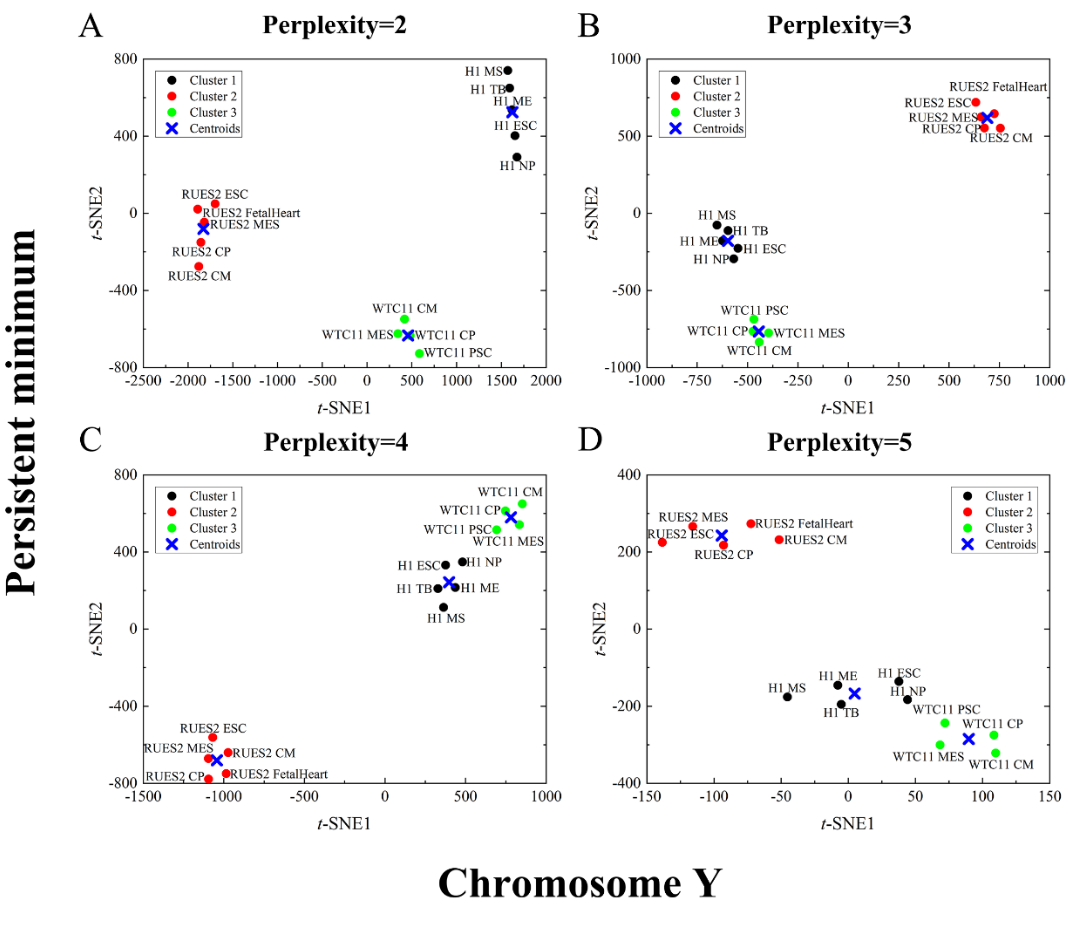

Figure S2. Performance comparison of PerSpectSC-ML's (here, ML method is *t*-SNE-assisted *k*-means) results obtained on persistent minimum of chromosome Y when the perplexity hyperparameters of *t*-SNE are 2 (**A**), 3 (**B**), 4 (**C**), and 5 (**D**) respectively.

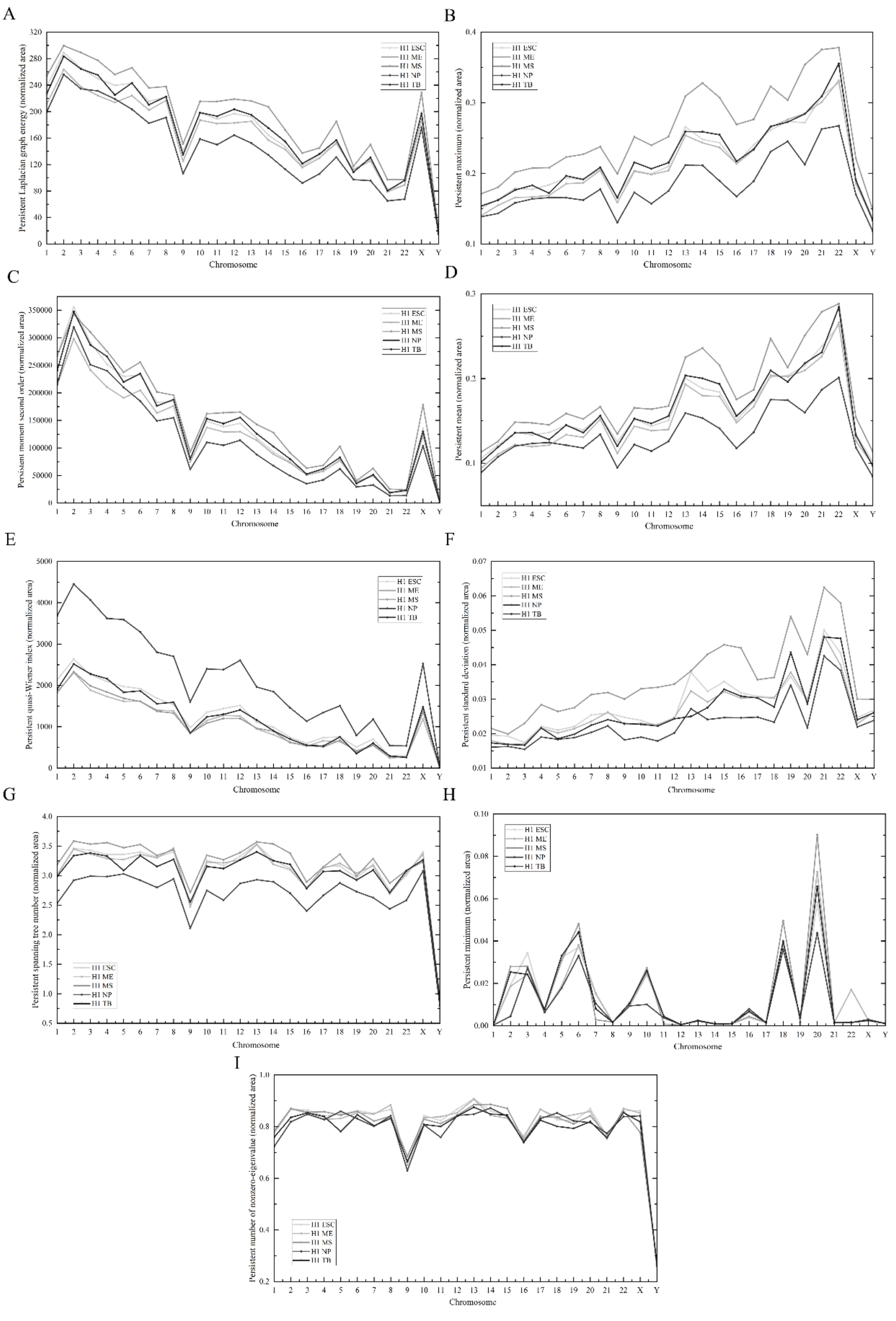

Figure S3. Performance comparison of normalized areas obtained from persistent Laplacian graph energy (**A**), persistent maximum (**B**), persistent moment second order (**C**), persistent mean (**D**), persistent quasi-Wiener index (**E**), persistent standard deviation (**F**), persistent spanning tree number (**G**), persistent minimum (**H**) and persistent number of nonzero-eigenvalue (**I**) of 24 chromosomes for H1 ESC, H1 ME, H1 MS, H1 NP and H1 TB.

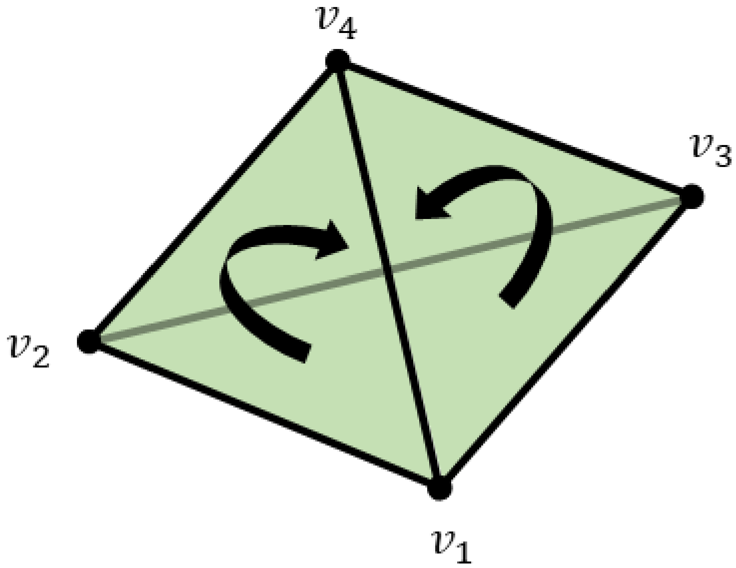

Figure S4. Illustration of the oriented simplex complex K_1.

We consider an oriented simplicial complex K_1 as in Figure S4. Its boundary operators are

$$\left\{\{v_1,v_2\}\ \{v_1,v_3\}\ \{v_1,v_4\}\ \{v_2,v_3\}\ \{v_2,v_4\}\ \{v_3,v_4\}\right\}_{\text{1-simplex}}$$

$$B_1=\begin{pmatrix} -1 & -1 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & -1 & 0 \\ 0 & 1 & 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}\left\{\begin{matrix}\{v_1\}\\ \{v_2\}\\ \{v_3\}\\ \{v_4\}\end{matrix}\right\}_{0\text{-simplex}} \quad \text{and}$$

$$\left\{\{v_1,v_2,v_3\}\ \{v_1,v_2,v_4\}\ \{v_1,v_3,v_4\}\ \{v_2,v_3,v_4\}\right\}_{\text{2-simplex}}$$

$$B_2=\begin{pmatrix} 1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & -1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 1 \end{pmatrix}\left\{\begin{matrix}\{v_1,v_2\}\\ \{v_1,v_3\}\\ \{v_1,v_4\}\\ \{v_2,v_3\}\\ \{v_2,v_4\}\\ \{v_3,v_4\}\end{matrix}\right\}_{1\text{-simplex}}$$

Here, the 0-simplex is denoted as $\{v_i\}$, 1-simplex is denoted as $\{v_i,\ v_j\}$ and 2-simplex is denoted as $\{v_i,\ v_j,\ v_k\}$. The corresponding Hodge Laplacian matrices are as follows:

$$\{\{v_1\}\ \{v_2\}\ \{v_3\}\ \{v_4\}\}_{\text{0-simplex}}$$

$$L_0 = \begin{pmatrix} 3 & -1 & -1 & -1 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 3 & -1 \\ -1 & -1 & -1 & 3 \end{pmatrix} \begin{Bmatrix} \{v_1\} \\ \{v_2\} \\ \{v_3\} \\ \{v_4\} \end{Bmatrix}_{\text{0-simplex}}$$

and

$$\{\{v_1, v_2\}\ \{v_1, v_3\}\ \{v_1, v_4\}\ \{v_2, v_3\}\ \{v_2, v_4\}\ \{v_3, v_4\}\}_{\text{1-simplex}}$$

$$L_1 = \begin{pmatrix} 4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 4 \end{pmatrix} \begin{Bmatrix} \{v_1, v_2\} \\ \{v_1, v_3\} \\ \{v_1, v_4\} \\ \{v_2, v_3\} \\ \{v_2, v_4\} \\ \{v_3, v_4\} \end{Bmatrix}_{\text{1-simplex}}$$

Note that mathematically L_k represents the topological connections in terms of upper and lower adjacency between k-simplexes.

Table S1. Hyperparameter settings for *t*-SNE

Hyperparameter	**Value**
Algorithm	exact
Distance	cosine
Exaggeration	4
NumDimensions	2
NumPCAComponets	0
Perplexity	6
Standardize	false
InitialY	1e-4*randn(N, NumDimensions)
LearnRate	500
NumPrint	20
MaxIter	1000
OutputFcn	[]
TolFun	1e-10
Theta	0.5
Verbose	0