Non-centered PCA

We formulated each map as a column of a data matrix:

$$\boldsymbol{X}_{g\times n} = \left(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\right) = \left(\boldsymbol{e}_1, \cdots, \boldsymbol{e}_g\right) \cdot \left[\begin{array}{c|c|c} x_{1,1} & \cdots & x_{1,n} \\ \vdots & \vdots & \vdots \\ x_{g,1} & \cdots & x_{g,n} \end{array}\right],$$

where g is the vector length and n is the number of cell types; $\boldsymbol{e}_i$ is the standard basis of $\boldsymbol{R}^g$. Assuming $\boldsymbol{A} = (a_{i,j})_{n\times n} = (\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n)$ where $\boldsymbol{a}_i$ is orthonormal eigenvector of $\boldsymbol{X}^T\boldsymbol{X}$ with eigenvalue λ_i, then $\boldsymbol{A}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{A} = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$. Let $\boldsymbol{F} = (\boldsymbol{f}_1, \cdots, \boldsymbol{f}_n) = \boldsymbol{X}\boldsymbol{A}$, then $\boldsymbol{X} = \boldsymbol{F}\boldsymbol{A}^T$, that is:

$$\boldsymbol{X} = \left(\boldsymbol{PC}_1, \cdots, \boldsymbol{PC}_n\right) \cdot \left(\begin{array}{c|c|c} a_{1,1}\cdot\sqrt{\lambda_1} & \cdots & a_{n,1}\cdot\sqrt{\lambda_1} \\ \vdots & \vdots & \vdots \\ \hline a_{1,n}\cdot\sqrt{\lambda_n} & \cdots & a_{n,n}\cdot\sqrt{\lambda_n} \end{array}\right),$$

where $\boldsymbol{PC}_i = \boldsymbol{f}_i \big/ \sqrt{\lambda_i}$ are orthonormal basis of $\boldsymbol{R}^n$.

The above formula enables the chromatin maps to be projected into an n-dimentional space spanned by $\boldsymbol{PC}_1, \cdots, \boldsymbol{PC}_n$. The variance along $\boldsymbol{PC}_i$ is $Var\left(a_{1,i}\cdot\sqrt{\lambda_i}, \cdots, a_{n,i}\cdot\sqrt{\lambda_i}\right) = (1/n - \bar{\boldsymbol{a}}_i^2)\cdot\lambda_i$, and the variation among columns of $\boldsymbol{X}$ is the sum of variance along each PC:

$$Variation(\boldsymbol{X}) = \frac{1}{n}\cdot\sum_{i=1}^{n}\left\|\boldsymbol{x}_i - \bar{\boldsymbol{x}}_i\right\|^2 = \sum_{i=1}^{n}(1/n - \bar{\boldsymbol{a}}_i^2)\cdot\lambda_i \ .$$

We sorted $\boldsymbol{PC}_i$ by variance and used the first three PCs for mapping which capture the largest fraction of variation. PCA requires calculating the eigenvalues and eigenvectors of $\boldsymbol{X}^T\boldsymbol{X}$. By standardizing chromatin map $\boldsymbol{x}_j$ as $x'_{ij} = \dfrac{x_{ij} - \bar{\boldsymbol{x}}_j}{\sqrt{\sum_{i=1}^{g}(x_{ij} - \bar{\boldsymbol{x}}_j)^2}}$, the $\boldsymbol{X}'^T\boldsymbol{X}'$ is equal to the correlation matrix of $\boldsymbol{X}$.