Supplementary Materials

Theorem S1. *Fan (1949) Let* $\mathbf{K}$ *is a symmetric matrix where* $\mathbf{u}_1, \cdots, \mathbf{u}_n$ *are eigenvectors corresponding to eigenvalues* $\lambda_1 \geq \cdots \geq \lambda_n$ *of* $\mathbf{K}$*. Then, the optimal solution of the problem*

$$\underset{\mathbf{H} \in \mathcal{R}^{n \times k}}{\operatorname{argmax}} \boldsymbol{tr}\left(\mathbf{H}^T \mathbf{K} \mathbf{H}\right)$$
$$\textit{subject to } \mathbf{H}^T \mathbf{H} = \mathbf{I}_k$$

is given by $\mathbf{H}^* = \mathbf{U}_k \mathbf{Q}$ *where* $\mathbf{U}_k = [\mathbf{u}_1, \cdots, \mathbf{u}_k]$ *and* Q *is an arbitrary* $k \times k$ *orthogonal matrix, and the maximum is given by*

$$\max_{\mathbf{H} \in \mathcal{R}^{n \times k}} \boldsymbol{tr}\left(\mathbf{H}^T \mathbf{K} \mathbf{H}\right) = \boldsymbol{tr}\left(\mathbf{H}^{*T} \mathbf{K} \mathbf{H}^*\right) = \sum_{i=1}^{k} \lambda_i$$

Proposition S1. *Suppose that* $\mathbf{X}^{(v)}$ *is a* $n \times p_v$ *centered data matrix from* n *samples and* p_v *random variables, that* $\mathbf{X} = \left[\mathbf{X}^{(1)}, \cdots, \mathbf{X}^{(m)}\right]$ *is an* $n \times p$ *multiview data matrix collected from* m *multiple sources where* $p = p_1 + \cdots + p_m$*, and that* $\mathbf{K}^{(v)} = \mathbf{X}^{(v)}\mathbf{X}^{(v)^T}$*. Then,*

$$\boldsymbol{tr}\left(\mathbf{H}^T\mathbf{K}^{(v)}\mathbf{H}\right) = \boldsymbol{tr}\left(\mathbf{V}_{1:k}^{(v)^T}\mathbf{X}^{(v)^T}\mathbf{X}^{(v)}\mathbf{V}_{1:k}^{(v)}\right) + \sum_{w \neq v} \boldsymbol{tr}\left(\mathbf{V}_{1:k}^{(v)^T}\mathbf{X}^{(v)^T}\mathbf{X}^{(w)}\mathbf{V}_{1:k}^{(w)}\right)$$

where $\mathbf{H} = \mathbf{U}_k\mathbf{Q}$ *is a* $n \times k$ *matrix where the column of* $\mathbf{U}_k$ *contains the first* k *eigenvectors of* $\mathbf{X}\mathbf{X}^T$ *corresponding to* k *largest eigenvalues,* $\mathbf{Q}$ *is an arbitrary orthogonal matrix, and* $\mathbf{V}_{1:k}^{(v)^T}$ *is a* $k \times p_v$ *matrix including the top* k *rows of* $\mathbf{V}^{(v)^T}$ *where* $\mathbf{V}^{(1)^T}$ *is a* $p \times p_1$ *matrix including the first* p_1 *eigenvector of* $\mathbf{X}^T\mathbf{X}$*,* $\mathbf{V}^{(2)^T}$ *is a* $p \times p_2$ *matrix including the next* p_2 *eigenvector of* $\mathbf{X}^T\mathbf{X}$ *and so on.*

Proof.

We consider the singular value decomposition of the $n \times p$ matrix $\mathbf{X}$:

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^T$$

where $\mathbf{U}$ is an $n \times n$ orthogonal matrix whose columns are the the left-singular vectors of $\mathbf{X}$, $\mathbf{D}$ is an $n \times p$ rectangular diagonal matrix with non-negative real values $\sigma_1, \cdots, \sigma_p$ known as singular values of $\mathbf{X}$ on the diagonal, and $\mathbf{V}$ is an $p \times p$ orthogonal matrix whose columns are the right-singular vectors of $\mathbf{X}$. Without loss of generality, we assume the singular values of $\mathbf{X}$, $\sigma_1 \geq \cdots \geq \sigma_p$, are ordered by largest to smallest and so the corresponding column vectors of $\mathbf{U}$ and $\mathbf{V}$ are as well.

Note that the c-th column vector of $\mathbf{V}$ (i.e. c-th right-singular vector of $\mathbf{X}$) is equivalent to the c-th eigenvector corresponding to the c-th largest eigenvalue λ_c of $\mathbf{X}^T\mathbf{X}$. Then we get:

$$\begin{aligned}
& \mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^T \\
\Leftrightarrow \quad & \left[\mathbf{X}^{(1)}, \cdots, \mathbf{X}^{(m)}\right] = \mathbf{U}\mathbf{D}\left[\mathbf{V}^{(1)^T}, \cdots, \mathbf{V}^{(m)^T}\right] \\
\Leftrightarrow \quad & \mathbf{X}^{(v)} = \mathbf{U}\mathbf{D}\mathbf{V}^{(v)^T} \qquad \text{for } v = 1, \cdots, m \\
\Leftrightarrow \quad & \mathbf{U}^T\mathbf{X}^{(v)} = \mathbf{D}\mathbf{V}^{(v)^T} \qquad \text{for } v = 1, \cdots, m
\end{aligned} \tag{S1}$$

Note that the c-th column vector of $\mathbf{U}$ (i.e. c-th left-singular vector of $\mathbf{X}$) is equivalent to the c-th eigenvector corresponding to the c-th largest eigenvalue λ_c of $\mathbf{X}\mathbf{X}^T$. Hence, $\mathbf{U} = [\mathbf{U}_k, \mathbf{U}_C]$ where the columns of $\mathbf{U}_k$ and $\mathbf{U}_C$ are the first k and the last $n - k$ eigenvectors of $\mathbf{X}\mathbf{X}^T$ respectively.

From (S1), we get

$$\mathbf{U}_k^T\mathbf{X}^{(v)} = \mathbf{D}_k\mathbf{V}_{1:k}^{(v)^T} \qquad \text{for } v = 1, \cdots, m \tag{S2}$$

where $\mathbf{D}_k$ is a $k \times k$ diagonal matrix with the first k singular values $\sigma_1, \cdots, \sigma_k$ on the diagonal and $\mathbf{V}_{1:k}^{(v)^T}$ is a $k \times p_v$ matrix including the top k rows of $\mathbf{V}^{(v)^T}$.

Using (S2), we get

$$\begin{aligned}
\mathbf{H}^T\mathbf{K}^{(v)}\mathbf{H} &= \mathbf{H}^T\mathbf{X}^{(v)}\mathbf{X}^{(v)^T}\mathbf{H} \\
&= \mathbf{Q}^T\mathbf{U}_k^T\mathbf{X}^{(v)}\mathbf{X}^{(v)^T}\mathbf{U}_k\mathbf{Q} \\
&= \mathbf{Q}^T\mathbf{D}_k\mathbf{V}_{1:k}^{(v)^T}\mathbf{V}_{1:k}^{(v)}\mathbf{D}_k^T\mathbf{Q}
\end{aligned}$$

Therefore,

$$\mathbf{tr}\left(\mathbf{H}^T\mathbf{K}^{(v)}\mathbf{H}\right) = \mathbf{tr}\left(\mathbf{Q}^T\mathbf{D}_k{\mathbf{V}_{1:k}^{(v)}}^T\mathbf{V}_{1:k}^{(v)}\mathbf{D}_k^T\mathbf{Q}\right)$$

By the invariant property under cyclic permutations of the trace function and the fact that $\mathbf{Q}$ is an orthogonal matrix, we get

$$\begin{aligned}\mathbf{tr}\left(\mathbf{H}^T\mathbf{K}^{(v)}\mathbf{H}\right) &= \mathbf{tr}\left(\mathbf{V}_{1:k}^{(v)}\mathbf{D}_k^T\mathbf{Q}\mathbf{Q}^T\mathbf{D}_k{\mathbf{V}_{1:k}^{(v)}}^T\right)\\ &= \mathbf{tr}\left(\mathbf{V}_{1:k}^{(v)}\mathbf{D}_k^T\mathbf{D}_k{\mathbf{V}_{1:k}^{(v)}}^T\right) \qquad \text{(S3)}\end{aligned}$$

Note that the c-th column vector of $\mathbf{V}$ (i.e. c-th right-singular vector of $\mathbf{X}$) is equivalent to the c-th eigenvector corresponding to the c-th largest eigenvalue λ_c of $\mathbf{X}^T\mathbf{X}$ (i.e. $\sigma_c^2 = \lambda_c$). Therefore, we have:

$$\mathbf{X}^T\mathbf{X}\mathbf{v}_c = \lambda_c\mathbf{v}_c$$

for $c = 1, \cdots, p$. Note that the c-th diagonal element of $\mathbf{D}$ is equivalent to the square roots of the c-th eigenvalue of $\mathbf{X}^T\mathbf{X}$. Therefore, we have:

$$\begin{aligned}&\mathbf{X}^T\mathbf{X}\mathbf{V}_{1:k} = \mathbf{V}_{1:k}\mathbf{D}_k^T\mathbf{D}_k\\ \Rightarrow\ &\mathbf{X}^T\mathbf{X}\mathbf{V}_{1:k}\mathbf{V}_{1:k}^T = \mathbf{V}_{1:k}\mathbf{D}_{1:k}^T\mathbf{D}_{1:k}\mathbf{V}_{1:k}^T \qquad \text{(S4)}\end{aligned}$$

The light-hand-side of (S4) has $p_v \times p_v$ square matrices (blocks) in the main diagonal as follow:

$$\begin{aligned}\mathbf{V}_{1:k}\mathbf{D}_{1:k}^T\mathbf{D}_{1:k}\mathbf{V}_{1:k}^T &= \left[\mathbf{D}_k{\mathbf{V}_{1:k}^{(1)}}^T, \cdots, \mathbf{D}_k{\mathbf{V}_{1:k}^{(m)}}^T\right]^T\left[\mathbf{D}_k{\mathbf{V}_{1:k}^{(1)}}^T, \cdots, \mathbf{D}_k{\mathbf{V}_{1:k}^{(m)}}^T\right]\\ &= \begin{bmatrix}\mathbf{V}_{1:k}^{(1)}\mathbf{D}_k^T\mathbf{D}_k{\mathbf{V}_{1:k}^{(1)}}^T & & \text{off-diagonal entries}\\ & \ddots & \\ \text{off-diagonal entries} & & \mathbf{V}_{1:k}^{(m)}\mathbf{D}_k^T\mathbf{D}_k{\mathbf{V}_{1:k}^{(m)}}^T\end{bmatrix}\end{aligned}$$

The left-hand-side of (S4) has $p_v \times p_v$ square matrices (blocks) in the main diagonal as follow:

$$\begin{aligned}&\mathbf{X}^T\mathbf{X}\mathbf{V}_{1:k}\mathbf{V}_{1:k}^T\\ &= \begin{bmatrix}\mathbf{X}^{(1)^T}\mathbf{X}^{(1)}\mathbf{V}_{1:k}^{(1)}{\mathbf{V}_{1:k}^{(1)}}^T + \sum_{w\neq 1}\mathbf{X}^{(1)^T}\mathbf{X}^{(w)}\mathbf{V}_{1:k}^{(w)}{\mathbf{V}_{1:k}^{(1)}}^T & & \text{off-diagonal entries}\\ & \ddots & \\ \text{off-diagonal entries} & & \mathbf{X}^{(m)^T}\mathbf{X}^{(m)}\mathbf{V}_{1:k}^{(m)}{\mathbf{V}_{1:k}^{(m)}}^T + \sum_{w\neq m}\mathbf{X}^{(m)^T}\mathbf{X}^{(w)}\mathbf{V}_{1:k}^{(w)}{\mathbf{V}_{1:k}^{(m)}}^T\end{bmatrix}\end{aligned}$$

Therefore, we have

$$\begin{aligned}&\mathbf{X}^{(v)^T}\mathbf{X}^{(v)}\mathbf{V}_{1:k}^{(v)}{\mathbf{V}_{1:k}^{(v)}}^T + \sum_{w\neq v}\mathbf{X}^{(v)^T}\mathbf{X}^{(w)}\mathbf{V}_{1:k}^{(w)}{\mathbf{V}_{1:k}^{(v)}}^T = \mathbf{V}_{1:k}^{(v)}\mathbf{D}_k^T\mathbf{D}_k{\mathbf{V}_{1:k}^{(v)}}^T\\ \Rightarrow\ &\mathbf{tr}\left({\mathbf{V}_{1:k}^{(v)}}^T\mathbf{X}^{(v)^T}\mathbf{X}^{(v)}\mathbf{V}_{1:k}^{(v)}\right) + \sum_{w\neq v}\mathbf{tr}\left({\mathbf{V}_{1:k}^{(v)}}^T\mathbf{X}^{(v)^T}\mathbf{X}^{(w)}\mathbf{V}_{1:k}^{(w)}\right) = \mathbf{tr}\left(\mathbf{V}_{1:k}^{(v)}\mathbf{D}_k^T\mathbf{D}_k{\mathbf{V}_{1:k}^{(v)}}^T\right)\end{aligned}$$

for $v = 1, \cdots, m$. From (S3), that is:

$$\mathbf{tr}\left(\mathbf{H}^T\mathbf{K}^{(v)}\mathbf{H}\right) = \mathbf{tr}\left(\mathbf{V}_{1:k}^{(v)^T}\mathbf{X}^{(v)^T}\mathbf{X}^{(v)}\mathbf{V}_{1:k}^{(v)}\right) + \sum_{w\neq v}\mathbf{tr}\left(\mathbf{V}_{1:k}^{(v)^T}\mathbf{X}^{(v)^T}\mathbf{X}^{(w)}\mathbf{V}_{1:k}^{(w)}\right)$$

□

Proposition S2. *Suppose that* $\mathbf{X}^{(v)}$ *is a* $n \times p_v$ *centered data matrix from* n *samples and* p_v *random variables, that* $\mathbf{X} = \left[\mathbf{X}^{(1)}, \cdots, \mathbf{X}^{(m)}\right]$ *is an* $n \times p$ *multiview data matrix collected from* m *multiple sources where* $p = p_1 + \cdots + p_m$*, and that* $\mathbf{K}^{(v)} = \mathbf{X}^{(v)}\mathbf{X}^{(v)^T}$*. Then,*

$$\begin{aligned} \boldsymbol{tr}\left(\mathbf{K}^{(v)} - \mathbf{H}^T\mathbf{K}^{(v)}\mathbf{H}\right) = \ & \boldsymbol{tr}\left(\mathbf{X}^{(v)}\mathbf{X}^{(v)^T}\right) \\ & - \left(\boldsymbol{tr}\left(\mathbf{V}_{1:k}^{(v)^T}\mathbf{X}^{(v)^T}\mathbf{X}^{(v)}\mathbf{V}_{1:k}^{(v)}\right) + \sum_{w \neq v} \boldsymbol{tr}\left(\mathbf{V}_{1:k}^{(v)^T}\mathbf{X}^{(v)^T}\mathbf{X}^{(w)}\mathbf{V}_{1:k}^{(w)}\right)\right) \end{aligned}$$

where $\mathbf{H} = \mathbf{U}_k\mathbf{Q}$ *is a* $n \times k$ *matrix where the column of* $\mathbf{U}_k$ *contains the first* k *eigenvectors of* $\mathbf{X}\mathbf{X}^T$ *corresponding to* k *largest eigenvalues,* $\mathbf{Q}$ *is an arbitrary orthogonal matrix, and* $\mathbf{V}_{1:k}^{(v)^T}$ *is a* $k \times p_v$ *matrix including the top* k *rows of* $\mathbf{V}^{(v)^T}$ *where* $\mathbf{V}^{(1)^T}$ *is a* $p \times p_1$ *matrix including the first* p_1 *eigenvector of* $\mathbf{X}^T\mathbf{X}$*,* $\mathbf{V}^{(2)^T}$ *is a* $p \times p_2$ *matrix including the next* p_2 *eigenvector of* $\mathbf{X}^T\mathbf{X}$ *and so on.*

Hence, $\boldsymbol{tr}\left(\mathbf{K}^{(v)} - \mathbf{H}^T\mathbf{K}^{(v)}\mathbf{H}\right)$ *can be interpreted as the unexplained variability (both the unexplained variance of the view* v *and the unexplained covariance of the view* v *with the other views* w*) by principal components of the joint representation.*

Proof.

Suppose that $\mathbf{\Phi}^{(v)}$ is a $n \times d_v$ centered data matrix from n samples and d_v random variables in the nonlinear feature space $\mathcal{F}$, and that $\mathbf{\Phi} = \left[\mathbf{\Phi}^{(1)}, \cdots, \mathbf{\Phi}^{(v)}\right]$ is a $n \times d$ multiview data matrix collected from multiple sources $v = 1, \cdots, m$ where $d = d_1 + \cdots + d_m$ and m is the number of views.

By Proposition S1, we knows

$$\mathbf{tr}\left(\mathbf{H}^T\mathbf{K}^{(v)}\mathbf{H}\right) = \mathbf{tr}\left(\mathbf{V}_{1:k}^{(v)^T}\mathbf{X}^{(v)^T}\mathbf{X}^{(v)}\mathbf{V}_{1:k}^{(v)}\right) + \sum_{w \neq v} \mathbf{tr}\left(\mathbf{V}_{1:k}^{(v)^T}\mathbf{X}^{(v)^T}\mathbf{X}^{(w)}\mathbf{V}_{1:k}^{(w)}\right)$$

And hence,

$$\begin{aligned} \mathbf{tr}\left(\mathbf{K}^{(v)} - \mathbf{H}^T\mathbf{K}^{(v)}\mathbf{H}\right) = \ & \mathbf{tr}\left(\mathbf{X}^{(v)}\mathbf{X}^{(v)^T}\right) \\ & - \left(\mathbf{tr}\left(\mathbf{V}_{1:k}^{(v)^T}\mathbf{X}^{(v)^T}\mathbf{X}^{(v)}\mathbf{V}_{1:k}^{(v)}\right) + \sum_{w \neq v} \mathbf{tr}\left(\mathbf{V}_{1:k}^{(v)^T}\mathbf{X}^{(v)^T}\mathbf{X}^{(w)}\mathbf{V}_{1:k}^{(w)}\right)\right) \end{aligned}$$

where $\mathbf{tr}\left(\mathbf{X}^{(v)}\mathbf{X}^{(v)^T}\right)$ is the total variance of the view v; $\mathbf{tr}\left(\mathbf{V}_{1:k}^{(v)^T}\mathbf{X}^{(v)^T}\mathbf{X}^{(v)}\mathbf{V}_{1:k}^{(v)}\right)$ is the variance of the view v explained by the eigenvectors of the feature space; $\mathbf{tr}\left(\mathbf{V}_{1:k}^{(v)^T}\mathbf{X}^{(v)^T}\mathbf{X}^{(w)}\mathbf{V}_{1:k}^{(w)}\right)$ is the covariance of the view v with the view w explained by the eigenvectors of the feature space. □

Proposition S3. *The optimization problem*

$$\underset{\boldsymbol{\theta}}{\text{maximize}} \sum_{v=1}^{m} \theta^{(v)} \boldsymbol{tr}\left(\mathbf{K}^{(v)} - \mathbf{H}^T\mathbf{K}^{(v)}\mathbf{H}\right) \quad \text{subject to } \frac{1}{2}\boldsymbol{\theta}^T\mathbf{Q}_m\boldsymbol{\theta} \leq 1,\ \boldsymbol{\theta} \geq \mathbf{0} \tag{S5}$$

has a closed form solution

$$\boldsymbol{\theta} = \left(\frac{a^{(1)}}{\sqrt{\left(a^{(1)}\right)^2 + \cdots + \left(a^{(m)}\right)^2}}, \cdots, \frac{a^{(m)}}{\sqrt{\left(a^{(1)}\right)^2 + \cdots + \left(a^{(m)}\right)^2}}\right)$$

where $a^{(v)} = \boldsymbol{tr}\left(\mathbf{K}^{(v)} - \mathbf{H}^T\mathbf{K}^{(v)}\mathbf{H}\right)$ *for* $v = 1, \cdots, m$, $\mathbf{H}$ *is a real valued* $n \times k$ *matrix* $\mathbf{H}$ *such that* $\mathbf{H}^T\mathbf{H} = \mathbf{I}_k$, *and* $\mathbf{K}^{(v)}$ *is a* $n \times n$ *positive semidefinite matrix.*

Proof.

We solve the optimization problem with geometric perspective. First, we define an optimal plane:

$$k = \theta^{(1)}a^{(1)} + \cdots + \theta^{(m)}a^{(m)} \tag{S6}$$

The plane is restricted to one of those that touch or pass through the hypersphere $\left(\theta^{(1)}\right)^2 + \cdots + \left(\theta^{(m)}\right)^2 = 1$ where $\boldsymbol{\theta} \geq 0$ and $a^{(v)}$s have non-negative real values by Proposition S4. The optimal solution is obtained at which the plane maximizes $k > 0$. In order to obtain the tangent point, we first find the normal vector that is perpendicular to the surface of the plane and passes through the center of the hypersphere (i.e. $\boldsymbol{\theta} = \mathbf{0}$) as follow:

$$\frac{\theta^{(1)} - 0}{a^{(1)}} = \cdots = \frac{\theta^{(m)} - 0}{a^{(m)}} = t$$

which is equivalent to

$$\boldsymbol{\theta}(t) = \left(a^{(1)}t, \cdots, a^{(m)}t\right)$$

where t is a real valued scalar variable. The plane touches the hypersphere at which the normal vector passing through the surface of the hypersphere.

$$\begin{aligned} & \left(\theta^{(1)}\right)^2 + \cdots + \left(\theta^{(m)}\right)^2 = 1 \\ \Leftrightarrow\ & \left(a^{(1)}t\right)^2 + \cdots + \left(a^{(m)}t\right)^2 = 1 \\ \Leftrightarrow\ & \left(a^{(1)}\right)^2 + \cdots + \left(a^{(m)}\right)^2 = \frac{1}{t^2} \end{aligned}$$

(i) In order for the plane to be tangent to the hypersphere where $\boldsymbol{\theta} \geq 0$, t should be a positive real value, hence, we get

$$t = \frac{1}{\sqrt{\left(a^{(1)}\right)^2 + \cdots + \left(a^{(m)}\right)^2}}$$

(ii) In order for the plane to pass through the hypersphere where $\boldsymbol{\theta} \geq 0$, t should be a positive real value such that $\left(\theta^{(1)}\right)^2 + \cdots + \left(\theta^{(m)}\right)^2 < 1$. Therefore,

$$0 < t < \frac{1}{\sqrt{\left(a^{(1)}\right)^2 + \cdots + \left(a^{(m)}\right)^2}}$$

Note that for any $t_s < t_l$,

$$k(t_s) = \boldsymbol{\theta}(t_s)^T \mathbf{a} < \boldsymbol{\theta}(t_l)^T \mathbf{a} = k(t_l)$$

where $\mathbf{a} = \left[a^{(1)}, \cdots, a^{(m)}\right]^T$. Therefore, the plane has the maximum k when it touches the hypersphere and the tangent point on its surface is the optimal solution of $\boldsymbol{\theta}$. The tangent point where the plane touches the hypersphere is at $t = \left(\sqrt{\left(a^{(1)}\right)^2 + \cdots + \left(a^{(m)}\right)^2}\right)^{-1}$, hence, the tangent point is:

$$\boldsymbol{\theta} = \left(\frac{a^{(1)}}{\sqrt{\left(a^{(1)}\right)^2 + \cdots + \left(a^{(m)}\right)^2}}, \cdots, \frac{a^{(m)}}{\sqrt{\left(a^{(1)}\right)^2 + \cdots + \left(a^{(m)}\right)^2}}\right)$$

which will be the optimal solution of the optimization problem S3. □

Proposition S4. *If* $\mathbf{K}^{(v)}$ *is a* $n \times n$ *positive semidefinite matrix,* $\boldsymbol{tr}\left(\mathbf{K}^{(v)} - \mathbf{H}^T\mathbf{K}^{(v)}\mathbf{H}\right)$ *is non-negative for any real valued* $n \times k$ *matrix* $\mathbf{H}$ *such that* $\mathbf{H}^T\mathbf{H} = \mathbf{I}_k$

Proof.

From Theorem S1,

$$\begin{aligned}\min_{\mathbf{H}\in\mathcal{R}^{n\times k},\mathbf{H}^T\mathbf{H}=\mathbf{I}_k} \mathbf{tr}\left(\mathbf{K}^{(v)} - \mathbf{H}^T\mathbf{K}^{(v)}\mathbf{H}\right) &= \mathbf{tr}\left(\mathbf{K}^{(v)}\right) - \max_{\mathbf{H}\in\mathcal{R}^{n\times k},\mathbf{H}^T\mathbf{H}=\mathbf{I}_k} \mathbf{tr}\left(\mathbf{H}^T\mathbf{K}^{(v)}\mathbf{H}\right) \\ &= \sum_{i=1}^{n}\lambda_i - \sum_{i=1}^{k}\lambda_i \\ &= \sum_{i=k+1}^{n}\lambda_i\end{aligned}$$

Since $\mathbf{K}^{(v)}$ is positive semidefinite, all its eigenvalues are non-negative. Therefore,

$$\min_{\mathbf{H}\in\mathcal{R}^{n\times k},\mathbf{H}^T\mathbf{H}=\mathbf{I}_k} \mathbf{tr}\left(\mathbf{K}^{(v)} - \mathbf{H}^T\mathbf{K}^{(v)}\mathbf{H}\right) = \sum_{i=k+1}^{n}\lambda_i \geq 0$$

Finally, for any real valued $n \times k$ matrix $\mathbf{H}$ such that $\mathbf{H}^T\mathbf{H} = \mathbf{I}_k$,

$$\mathbf{tr}\left(\mathbf{K}^{(v)} - \mathbf{H}^T\mathbf{K}^{(v)}\mathbf{H}\right) \geq \min_{\mathbf{H}\in\mathcal{R}^{n\times k},\mathbf{H}^T\mathbf{H}=\mathbf{I}_k} \mathbf{tr}\left(\mathbf{K}^{(v)} - \mathbf{H}^T\mathbf{K}^{(v)}\mathbf{H}\right) \geq 0$$

□

Text S1. ***Centering and scaling.***

At every iteration, we must center the combined map $\boldsymbol{\phi}_{\boldsymbol{\theta}}(\mathbf{x}_i)$ around the origin before we perform (kernel) PCA and update the cluster assignments $\mathbf{H}$. That is, at every iteration, we must center the data by using the following kernel trick: $\mathbf{K}_{\boldsymbol{\theta}} \leftarrow \mathbf{K}_{\boldsymbol{\theta}} - \mathbf{J}_n\mathbf{K}_{\boldsymbol{\theta}} - \mathbf{K}_{\boldsymbol{\theta}}\mathbf{J}_n + \mathbf{J}_n\mathbf{K}_{\boldsymbol{\theta}}\mathbf{J}_n$ where $\mathbf{J}_n = \mathbf{1}_n\mathbf{1}_n^T/n$ (Schölkopf *et al.*, 1998). This is computationally inefficient. Therefore, we suggest the following proposition. Using this proposition, we center $\mathbf{K}^{(v)}$ for each view only at the beginning of the algorithm instead of centering the combined kernel matrix $\mathbf{K}_{\boldsymbol{\theta}}$ at every iterations.

Proposition S5. *Let* $\widetilde{\mathbf{K}}^*_{\boldsymbol{\theta}} = \sum_{v=1}^m \theta^{(v)}\widetilde{\mathbf{K}}^{(v)}$ *where* $\widetilde{\mathbf{K}}^{(v)} = \mathbf{K}^{(v)} - \mathbf{J}_n\mathbf{K}^{(v)} - \mathbf{K}^{(v)}\mathbf{J}_n + \mathbf{J}_n\mathbf{K}^{(v)}\mathbf{J}_n$ *for* $v = 1, \cdots, m$. *Then* $\widetilde{\mathbf{K}}^*_{\boldsymbol{\theta}} = \widetilde{\mathbf{K}}_{\boldsymbol{\theta}}$ *where* $\widetilde{\mathbf{K}}_{\boldsymbol{\theta}} = \mathbf{K}_{\boldsymbol{\theta}} - \mathbf{J}_n\mathbf{K}_{\boldsymbol{\theta}} - \mathbf{K}_{\boldsymbol{\theta}}\mathbf{J}_n + \mathbf{J}_n\mathbf{K}_{\boldsymbol{\theta}}\mathbf{J}_n$ *for any* $\boldsymbol{\theta} \in \mathcal{R}^m$.

Proof.

$$
\begin{aligned}
\widetilde{\mathbf{K}_{\boldsymbol{\theta}}}^* &= \sum_{v=1}^m \theta^{(v)}\widetilde{\mathbf{K}}^{(v)} \\
&= \sum_{v=1}^m \theta^{(v)}\left(\mathbf{K}^{(v)} - \mathbf{J}_n\mathbf{K}^{(v)} - \mathbf{K}^{(v)}\mathbf{J}_n + \mathbf{J}_n\mathbf{K}^{(v)}\mathbf{J}_n\right) \\
&= \sum_{v=1}^m \theta^{(v)}\mathbf{K}^{(v)} - \sum_{v=1}^m \theta^{(v)}\mathbf{J}_n\mathbf{K}^{(v)} - \sum_{v=1}^m \theta^{(v)}\mathbf{K}^{(v)}\mathbf{J}_n + \sum_{v=1}^m \theta^{(v)}\mathbf{J}_n\mathbf{K}^{(v)}\mathbf{J}_n \\
&= \mathbf{K}_{\boldsymbol{\theta}} - \mathbf{J}_n\left(\sum_{v=1}^m \theta^{(v)}\mathbf{K}^{(v)}\right) - \left(\sum_{v=1}^m \theta^{(v)}\mathbf{K}^{(v)}\right)\mathbf{J}_n + \mathbf{J}_n\left(\sum_{v=1}^m \theta^{(v)}\mathbf{K}^{(v)}\right)\mathbf{J}_n \\
&= \mathbf{K}_{\boldsymbol{\theta}} - \mathbf{J}_n\mathbf{K}_{\boldsymbol{\theta}} - \mathbf{K}_{\boldsymbol{\theta}}\mathbf{J}_n + \mathbf{J}_n\mathbf{K}_{\boldsymbol{\theta}}\mathbf{J}_n \\
&= \widetilde{\mathbf{K}_{\boldsymbol{\theta}}}
\end{aligned}
$$

□

It is known that estimation of kernel coefficients depends on how the kernel matrices are scaled (Kloft *et al.*, 2011; Ong and Zien, 2008). In order to make multiple views comparable to each other, we suggest to scale each kernel matrix before combining them by $\mathbf{K}^{(v)} \leftarrow \mathbf{K}^{(v)}/\mathbf{tr}\left(\mathbf{K}^{(v)}\right)$. Note that the trace of the centered kernel matrix is the sum of its eigenvalues, i.e. $\mathbf{tr}\left(\mathbf{K}^{(v)}\right) = \sum_{i=1}^n \lambda_i^{(v)}$, which can be interpreted as the measure of variance explained by principal components of the feature space $\mathcal{F}$ within each view. Therefore, by scaling the kernel matrix, the total variance explained within each view is set to be uniform, i.e. $\mathbf{tr}\left(\mathbf{K}^{(1)}\right) = \cdots = \mathbf{tr}\left(\mathbf{K}^{(m)}\right) = 1$.

Text S2. ***Simulation Detail.***

We evaluate robustness of our method against two types of adversarial perturbations:

- Noise variables that are independently sampled from Gaussian distribution with zero-mean and unit-variance. We add different numbers ($N_{noise} = 0, 1, 2, \cdots$) of noise variables to a view.
- Redundant variables that are correlated with original variables. We add different numbers ($N_{redun} = 1, 2, \cdots$) of variables having different correlations ($cor = 1, 0.97, 0.90, 0.72, 0.45$) with the original variables to a view.

Under these perturbations, we examine how our method make use of complementary patterns in multiple views. For this purpose, we first generated multiview data in three scenarios A–C with two or three views. Those views have complementary patterns necessary for identifying true clusters. Scenario A is composed of a complete view that has complete information to detect the three clusters and a partial view that only conveys partial information. Scenario B is composed of two different partial views so that each view alone cannot completely detect the three clusters. Both scenarios A & B aim to test how the compared methods use the complementary information in two views. Scenario C is composed of two different partial views and a noise view. It aims to test further whether the methods robustly use complementary information from views even when one of the views contains only noise variables. Then, we added different types and levels of adversarial perturbations to one of the views. We denote the simulation data with the noise variables by A-Noise, B-Noise, and C-Noise, and the data with the redundant variables by A-Redun, B-Redun, and C-Redun.

All features were standardized so that they are centered around zero with standard deviations of one. A kernel function $\mathbf{k}(\mathbf{x}, \mathbf{y}) = \exp\left(-0.5||\mathbf{x} - \mathbf{y}||^2\right)$ was used for all the views. After obtaining continuous clustering indicator $\mathbf{H}^*$, we performed k-means clustering on the normalized $\mathbf{H}^*$ with 1000 random starts and reported the best result minimizing the objective function. We stopped the iteration if the stopping criteria $||\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-1}||_2 < 10^{-4}$ is met within 500 iterations.

We compared MML-MKKC with seven other methods: two baseline methods, four recently proposed MKKC methods, and one variant of MML-MKKC. In particular, we included the following baseline methods:

- **Single Best** uses the best view that minimizes the kernel k-means objective function (1).
- **Uniform Weight** equally assigns all the kernel coefficients $\boldsymbol{\theta}$ to all views. It takes the combined kernel $\mathbf{K}_{\boldsymbol{\theta}} = \sum_{v=1}^{m} \mathbf{K}^{(v)}/m$ as an input $\mathbf{K}$ in the problem (1).

The following three MKKC methods are similar in that they all combine multiple kernels as: $\mathbf{K}_{\boldsymbol{\theta}} = \sum_{v=1}^{m} {\theta^{(v)}}^2 \mathbf{K}^{(v)}$, with l_1 constraint on the kernel coefficients $\boldsymbol{\theta}$, and solve the problem (2) using the $\min_{\mathbf{H}}$-$\min_{\boldsymbol{\theta}}$ framework.

- **Gonen's MKK** (Gönen and Margolin, 2014)
- **Gonen's LMKK** (Gönen and Margolin, 2014): This localized multiple kernel k-means clustering method aims to capture sample-specific characteristic of multiple data sources by estimating sample-specific kernel coefficients.
- **Liu's MKK-MIR** (Liu *et al.*, 2016): This method characterizes the correlation of each pair of kernels by integrating a matrix-induced quadratic regularization into the objective function. The regularization parameter λ was set to 1 and the quadratic coefficient matrix $\mathbf{M}$ was defined as suggested by the paper.

The fourth MKKC method combines the multiple kernels in a different way:

- **Yu's OKKC** (Yu *et al.*, 2012): This method combines multiple views as $\mathbf{K}_{\boldsymbol{\theta}} = \sum_{v=1}^{m} \theta^{(v)} \mathbf{K}^{(v)}$ and uses l_p constraint on $\boldsymbol{\theta}$ where $p \geq 1$, and optimize the problem (2) using the $\max_{\mathbf{H}}$-$\max_{\boldsymbol{\theta}}$ framework. However, rather than minimizing $\mathbf{tr}\left(\mathbf{K}_{\boldsymbol{\theta}}\right) - \mathbf{tr}\left(\mathbf{H}^T \mathbf{K}_{\boldsymbol{\theta}} \mathbf{H}\right)$ as the general formula (2) does, it maximizes the objective function $\mathbf{tr}\left(\mathbf{H}^T \mathbf{K}_{\boldsymbol{\theta}} \mathbf{H}\right)$ so that it also leads to solutions favor assigning more weights to dominant views. The original algorithm iteratively optimizes the kernel coefficients $\boldsymbol{\theta}$ and discrete clustering assignment, which increases computational burden and costs more time. For a fair comparison, we updated the continuous cluster assignment $\mathbf{H}$ instead of retaining the discrete assignment at every iteration, and optimized it as QCLP, as proposed by all the other MKKC methods including ours.

Finally, we also include a variant of our method in the comparison:

- **MinMax-MinC** is the l_1-regularization version of our method MML-MKKC, which is included to examine the effect of l_2-regularization in our method on clustering. It uses the same $\min_{\mathbf{H}}$-$\max_{\boldsymbol{\theta}}$ formulation in the problem (3) as our method but with l_1 instead of l_2 constraint on $\boldsymbol{\theta}$. Additionally, it uses $\boldsymbol{\theta} \geq \boldsymbol{\theta}_{min}$ where $\boldsymbol{\theta}_{min} = 0.5/m\mathbf{1}$ to avoid a sparse trivial solution.

Text S3. ***Data preprocessing and strategy.***

The mRNA and methylation data are log-transformed. Variables have more than 5% missing values are excluded, otherwise imputed using KNNimpute (Hastie *et al.*, 1999). For each cancer, the top 100 features with largest median absolute deviation across the samples are used for each view. A radial basis function kernel is used for all views as suggested by To avoid the kernel matrices getting zero values due to a large number of features, we set the parameter of the radial basis function kernel as $\sigma = 1/(2 + p^2)$ where p is the number of features. The kernel matrices were centered and scaled as described in Section S1.

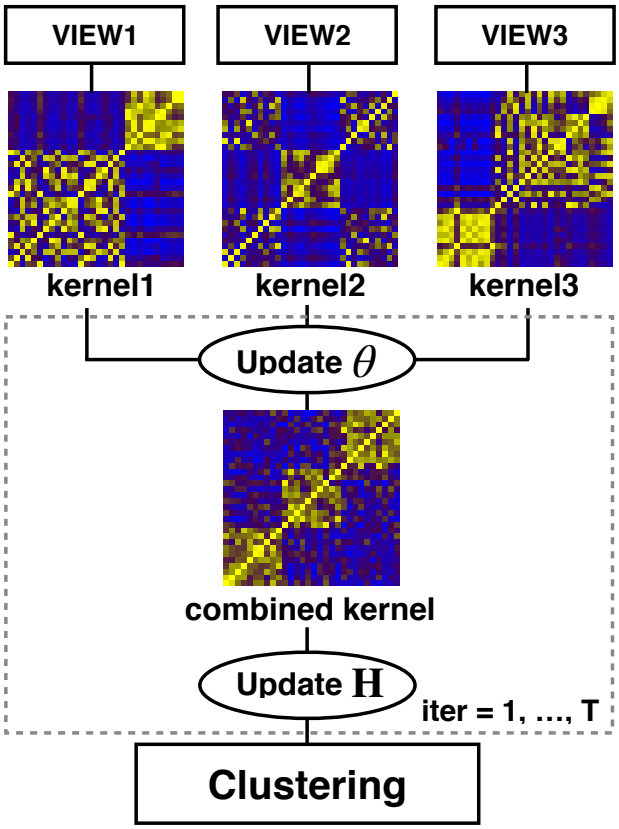

Figure S1: Overview of multiple kernel clustering. It combines multiple views by taking a linear sum of multiple kernels where each kernel captures similarity between samples within each view. The kernel coefficients $\boldsymbol{\theta}$ and the cluster assignment matrix $\mathbf{H}$ are alternately optimized given each other.

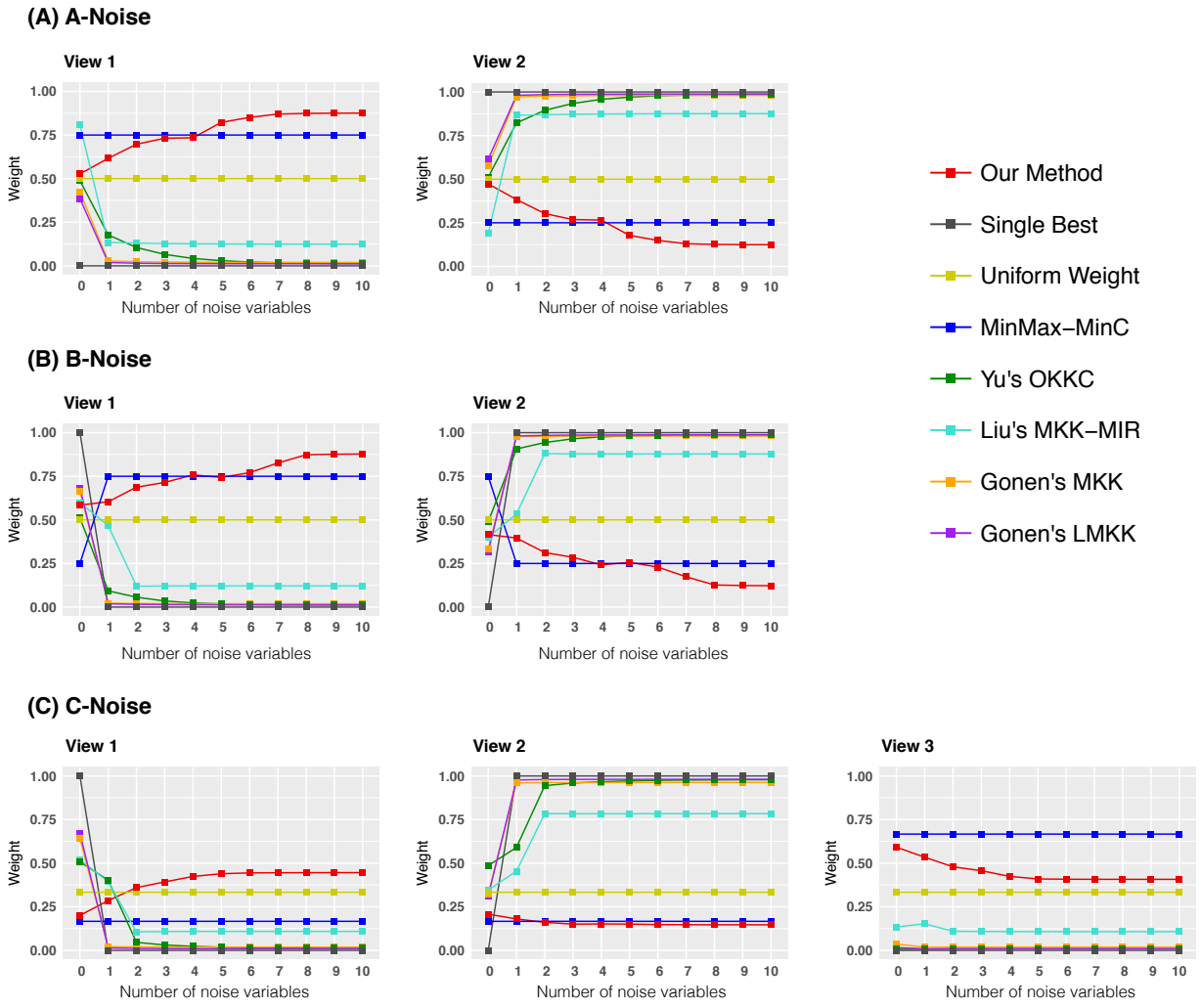

Figure S2: **Weights given by the compared methods to the views when Scenarios A-Noise, B-Noise, and C-Noise were used to identify clusters.** The weights on the views are plotted against the number of the noise variables $\mathbf{N}_{noise}$ added to the first view in each scenario. The x-axis represents the number of noise variables added to the first view. The y-axis represents the relative weight given by the compared methods. The methods are identified by different colors. For comparison purposes, we defined the weight as $\boldsymbol{\theta}/\boldsymbol{\theta}^T\mathbf{1}$ for the methods combining kernels using $\mathbf{K}_{\boldsymbol{\theta}} = \sum_{v=1}^{m} \theta^{(v)}\mathbf{K}^{(v)}$ (such as Uniform Weight, MinMax-MinC, Yu's OKKC, and our method) and as $\boldsymbol{\theta}^2/\boldsymbol{\theta}^{2T}\mathbf{1}$ for the methods using $\mathbf{K}_{\boldsymbol{\theta}} = \sum_{v=1}^{m} {\theta^{(v)}}^2\mathbf{K}^{(v)}$ (such as Liu's MIR, Gonen's MKK and LMKK).

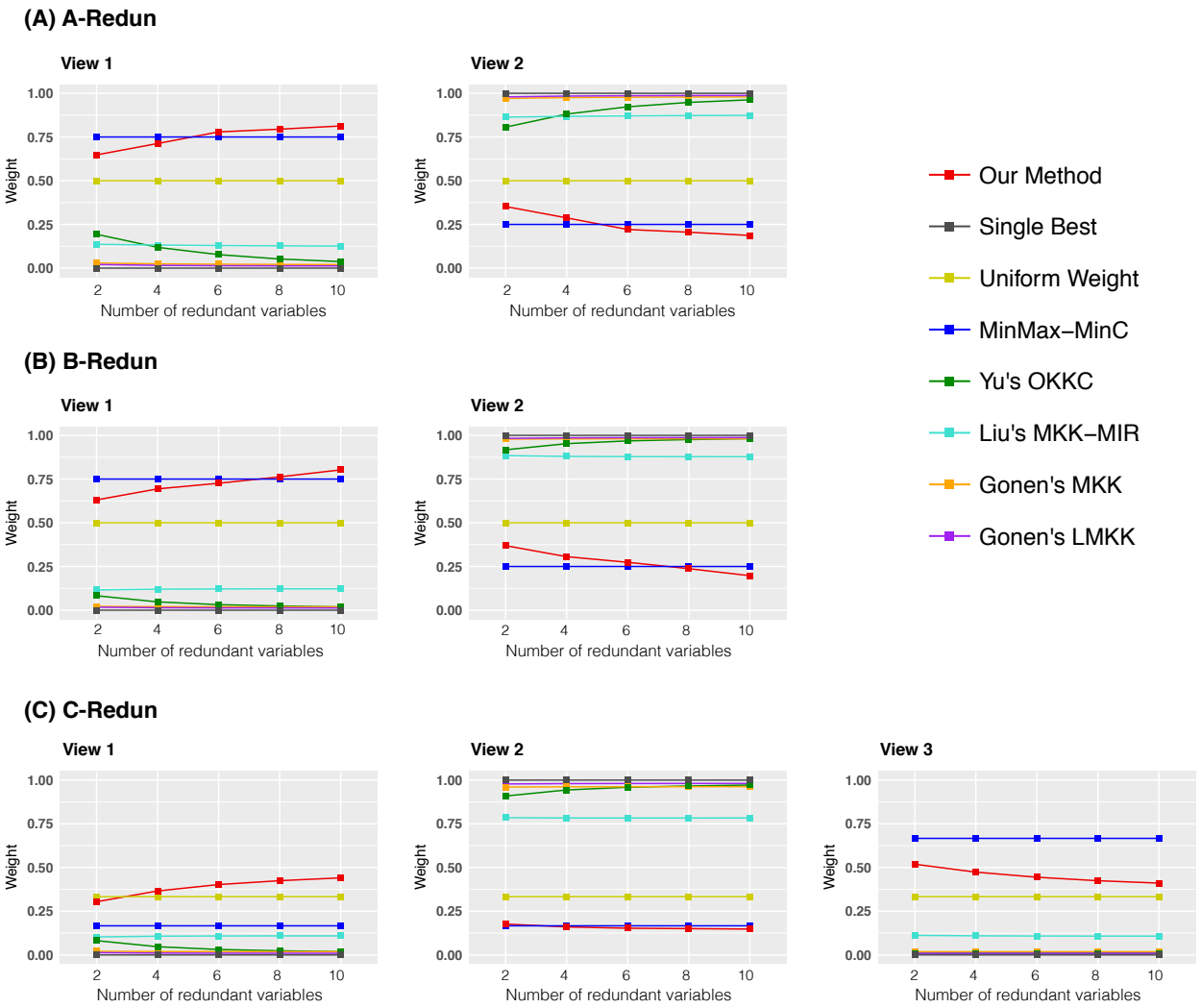

Figure S3: **Weights given by the compared methods to the views when Scenario A-2, B-2, and C-2 were used to identify clusters.** The weights on the views are plotted against the redundant variables $\mathbf{N}_{redun}$ where $cor = 0.90$ added to the first view in each scenario. The x-axis represents the number of noise variables added to the first view. The y-axis represents the relative weight given by the compared methods. The methods are identified by different colors. For comparison purposes, we defined the weight as $\boldsymbol{\theta}/\boldsymbol{\theta}^T\mathbf{1}$ for the methods combining kernels using $\mathbf{K}_{\boldsymbol{\theta}} = \sum_{v=1}^{m} \theta^{(v)}\mathbf{K}^{(v)}$ (such as Uniform Weight, MinMax-MinC, Yu's OKKC, and our method) and as $\boldsymbol{\theta}^2/\boldsymbol{\theta}^{2T}\mathbf{1}$ for the methods using $\mathbf{K}_{\boldsymbol{\theta}} = \sum_{v=1}^{m} {\theta^{(v)}}^2\mathbf{K}^{(v)}$ (such as Liu's MIR, Gonen's MKK and LMKK).

Table S1: **Evaluation of the Clutering Methods on Scenario A-Noise**

Scenario A-1												
	$\mathbf{N}_{noise}$	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Single Best	adjRI	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497
	normMI	0.837	0.837	0.837	0.837	0.837	0.837	0.837	0.837	0.837	0.837	**0.837**
	purity	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680
Uniform Weight	adjRI	**1.000**	**1.000**	0.795	0.522	0.502	0.500	0.499	0.499	0.498	0.497	**0.498**
	normMI	**1.443**	**1.443**	1.137	0.858	0.840	0.839	0.838	0.838	0.837	0.837	**0.837**
	purity	**1.000**	**1.000**	0.923	0.740	0.700	0.693	0.687	0.687	0.680	0.680	**0.687**
MinMax MinC	adjRI	**1.000**	**1.000**	0.795	0.522	0.502	0.500	0.499	0.499	0.498	0.497	**0.498**
	normMI	**1.443**	**1.443**	1.137	0.858	0.840	0.839	0.838	0.838	0.837	0.837	**0.837**
	purity	**1.000**	**1.000**	0.923	0.740	0.700	0.693	0.687	0.687	0.680	0.680	**0.687**
Yu's OKKC	adjRI	**1.000**	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497
	normMI	**1.443**	0.837	0.837	0.837	0.837	0.837	0.837	0.837	0.837	0.837	**0.837**
	purity	**1.000**	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680
Liu's MKK-MIR	adjRI	**1.000**	0.500	0.500	0.498	0.498	0.497	0.497	0.497	0.497	0.497	0.497
	normMI	**1.443**	0.839	0.839	0.837	0.837	0.837	0.837	0.837	0.837	0.837	**0.837**
	purity	**1.000**	0.693	0.693	0.683	0.683	0.680	0.680	0.680	0.680	0.680	0.680
Gonen's MKK	adjRI	**1.000**	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497
	normMI	**1.443**	0.837	0.837	0.837	0.837	0.837	0.837	0.837	0.837	0.837	**0.837**
	purity	**1.000**	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680
Gonen's LMKK	adjRI	0.980	0.498	0.498	0.498	0.498	0.498	0.498	0.498	0.498	0.498	**0.498**
	normMI	1.400	0.837	0.837	0.837	0.837	0.837	0.837	0.837	0.837	0.837	**0.837**
	purity	0.993	0.683	0.683	0.683	0.683	0.683	0.683	0.683	0.683	0.683	0.683
Our Method	adjRI	**1.000**	**1.000**	**1.000**	**0.951**	**0.666**	**0.649**	**0.548**	**0.508**	**0.503**	**0.501**	**0.498**
	normMI	**1.443**	**1.443**	**1.443**	**1.355**	**0.993**	**0.976**	**0.882**	**0.846**	**0.841**	**0.840**	**0.837**
	purity	**1.000**	**1.000**	**1.000**	**0.983**	**0.860**	**0.850**	**0.773**	**0.717**	**0.703**	**0.697**	0.683

Clustering performance of the methods are evaluated by three widely-used metrics: Adjusted Rand Index (adjRI), Normalized Mutual Information (normMI), and purity. A higher value of the metrics indicates better clustering performance. Each column represents a simulated data set where the corresponding number indicates the number of the noise variables ($\mathbf{N}_{noise}$) added to the complete view (View 1). The bolded numbers are the maximum value for each the evalution measure within a simulation data set.

Table S2: **Evaluation of the Clutering Methods on Scenario B-Noise**

Scenario B-1												
	$\mathbf{N}_{noise}$	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Single Best	adjRI	0.497	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	**0.487**	**0.487**
	normMI	0.837	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	**0.811**
	purity	0.667	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663
Uniform Weight	adjRI	**1.000**	**1.000**	0.980	0.487	0.487	0.487	0.487	0.487	0.487	**0.487**	**0.487**
	normMI	**1.443**	**1.443**	1.400	0.811	0.811	0.811	0.811	0.811	0.811	0.811	**0.811**
	purity	**1.000**	**1.000**	0.993	0.670	0.663	0.663	0.663	0.667	0.663	0.663	0.663
MinMax MinC	adjRI	**1.000**	**1.000**	0.980	0.487	0.487	0.487	0.487	0.487	0.487	**0.487**	**0.487**
	normMI	**1.443**	**1.443**	1.400	0.811	0.811	0.811	0.811	0.811	0.811	0.811	**0.811**
	purity	**1.000**	**1.000**	0.993	0.670	0.663	0.663	0.663	0.667	0.663	0.663	0.663
Yu's OKKC	adjRI	**1.000**	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	**0.487**	**0.487**
	normMI	**1.443**	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	**0.811**
	purity	**1.000**	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663
Liu's MKK-MIR	adjRI	**1.000**	**1.000**	0.487	0.487	0.487	0.487	0.487	0.487	0.487	**0.487**	**0.487**
	normMI	**1.443**	**1.443**	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	**0.811**
	purity	**1.000**	**1.000**	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663
Gonen's MKK	adjRI	**1.000**	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	**0.487**	**0.487**
	normMI	**1.443**	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	**0.811**
	purity	**1.000**	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663
Gonen's LMKK	adjRI	0.552	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	**0.487**	**0.487**
	normMI	0.832	0.812	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	**0.811**
	purity	0.820	0.670	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663
Our Method	adjRI	**1.000**	**1.000**	**1.000**	**0.951**	**0.789**	**0.510**	**0.490**	**0.493**	**0.490**	**0.487**	**0.487**
	normMI	**1.443**	**1.443**	**1.443**	**1.355**	**1.159**	**0.847**	**0.814**	**0.817**	**0.814**	**0.812**	**0.811**
	purity	**1.000**	**1.000**	**1.000**	**0.983**	**0.920**	**0.720**	**0.687**	**0.700**	**0.690**	**0.673**	**0.670**

Clustering performance of the methods are evaluated by three widely-used metrics: Adjusted Rand Index (adjRI), Normalized Mutual Information (normMI), and purity. A higher value of the metrics indicates better clustering performance. Each column represents a simulated data set where the corresponding number indicates the number of the noise variables ($\mathbf{N}_{noise}$) added to the first partial view (View 1). The bolded numbers are the maximum value for each the evalution measure within a simulation data set.

Table S3: **Evaluation of the Clutering Methods on Scenario C-Noise**

Scenario C-1												
	$\mathbf{N}_{noise}$	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Single Best	adjRI	0.497	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487
	normMI	0.837	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811
	purity	0.667	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663
Uniform Weight	adjRI	**1.000**	**0.990**	0.878	0.500	0.487	0.487	0.487	0.487	0.487	0.487	0.487
	normMI	**1.443**	**1.418**	1.234	0.839	0.811	0.812	0.811	0.811	0.812	0.812	0.812
	purity	**1.000**	**0.997**	0.957	0.693	0.667	0.670	0.667	0.667	0.667	0.667	0.667
MinMax MinC	adjRI	**1.000**	**0.990**	0.878	0.500	0.487	0.487	0.487	0.487	0.487	0.487	0.487
	normMI	**1.443**	**1.418**	1.234	0.839	0.811	0.812	0.811	0.811	0.812	0.812	0.812
	purity	**1.000**	**0.997**	0.957	0.693	0.667	0.670	0.667	0.667	0.667	0.667	0.667
Yu's OKKC	adjRI	**1.000**	0.545	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487
	normMI	**1.443**	0.879	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811
	purity	**1.000**	0.770	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663
Liu's MKK-MIR	adjRI	**1.000**	0.980	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487
	normMI	**1.443**	1.394	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811
	purity	**1.000**	0.993	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663
Gonen's MKK	adjRI	**1.000**	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487
	normMI	**1.443**	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811
	purity	**1.000**	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663
Gonen's LMKK	adjRI	0.500	0.490	0.490	0.490	0.490	0.490	0.490	0.490	0.490	0.490	0.490
	normMI	0.775	0.814	0.814	0.814	0.814	0.814	0.814	0.814	0.814	0.814	0.814
	purity	0.790	0.683	0.683	0.683	0.683	0.683	**0.683**	**0.683**	0.683	**0.683**	**0.683**
Our Method	adjRI	**1.000**	**0.990**	**0.923**	**0.869**	**0.556**	**0.515**	**0.498**	**0.498**	**0.499**	**0.498**	**0.497**
	normMI	**1.443**	**1.418**	**1.300**	**1.223**	**0.890**	**0.852**	**0.838**	**0.837**	**0.838**	**0.838**	**0.837**
	purity	**1.000**	**0.997**	**0.973**	**0.953**	**0.780**	**0.730**	0.680	0.680	**0.690**	**0.683**	0.677

Clustering performance of the methods are evaluated by three widely-used metrics: Adjusted Rand Index (adjRI), Normalized Mutual Information (normMI), and purity. A higher value of the metrics indicates better clustering performance. Each column represents a simulated data set where the corresponding number indicates the number of the noise variables ($\mathbf{N}_{noise}$) added to the first partial view (View 1). The bolded numbers are the maximum value for each the evalution measure within a simulation data set.

Table S4: **Evaluation of the Clutering Methods on Scenario A-Redun**

Scenario A-2																											
	cor	**0.45**					**0.72**					**0.90**					**0.97**					**1**				
	N$_{redun}$	**2**	**4**	**6**	**8**	**10**	**2**	**4**	**6**	**8**	**10**	**2**	**4**	**6**	**8**	**10**	**2**	**4**	**6**	**8**	**10**	**2**	**4**	**6**	**8**	**10**
Single Best	adjRI	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497
	normMI	0.837	0.837	**0.837**	**0.837**	**0.837**	0.837	0.837	0.837	0.837	**0.837**	0.837	0.837	0.837	0.837	0.837	0.837	0.837	0.837	0.837	0.837	0.837	0.837	0.837	0.837	0.837
	purity	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680
Uniform Weight	adjRI	**0.507**	0.497	0.497	0.497	0.497	0.795	0.499	0.497	0.497	0.497	**0.980**	0.951	0.515	0.501	0.499	**1.000**	**1.000**	**1.000**	**1.000**	0.904	**1.000**	**1.000**	**1.000**	**1.000**	**0.990**
	normMI	**0.845**	0.837	**0.837**	**0.837**	**0.837**	1.132	0.839	0.836	0.837	**0.837**	**1.400**	1.341	0.852	0.840	0.838	**1.443**	**1.443**	**1.443**	**1.443**	1.270	**1.443**	**1.443**	**1.443**	**1.443**	**1.418**
	purity	**0.713**	0.680	0.680	0.680	0.680	0.923	0.683	0.677	0.680	0.680	**0.993**	0.983	0.730	0.697	0.687	**1.000**	**1.000**	**1.000**	**1.000**	0.967	**1.000**	**1.000**	**1.000**	**1.000**	**0.997**
MinMax MinC	adjRI	**0.507**	0.497	0.497	0.497	0.497	0.795	0.499	0.497	0.497	0.497	**0.980**	0.951	0.515	0.501	0.499	**1.000**	**1.000**	**1.000**	**1.000**	0.904	**1.000**	**1.000**	**1.000**	**1.000**	**0.990**
	normMI	**0.845**	0.837	**0.837**	**0.837**	**0.837**	1.132	0.839	0.836	0.837	**0.837**	**1.400**	1.341	0.852	0.840	0.838	**1.443**	**1.443**	**1.443**	**1.443**	1.270	**1.443**	**1.443**	**1.443**	**1.443**	**1.418**
	purity	**0.713**	0.680	0.680	0.680	0.680	0.923	0.683	0.677	0.680	0.680	**0.993**	0.983	0.730	0.697	0.687	**1.000**	**1.000**	**1.000**	**1.000**	0.967	**1.000**	**1.000**	**1.000**	**1.000**	**0.997**
Yu's OKKC	adjRI	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	**1.000**	0.497	0.497	0.497	0.497	**1.000**	**1.000**	0.519	0.497	0.497
	normMI	0.837	0.837	**0.837**	**0.837**	**0.837**	0.837	0.837	0.837	0.837	**0.837**	0.837	0.837	0.837	0.837	0.837	**1.443**	0.837	0.837	0.837	0.837	**1.443**	**1.443**	0.856	0.837	0.837
	purity	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	**1.000**	0.680	0.680	0.680	0.680	**1.000**	**1.000**	0.737	0.680	0.680
Liu's MKK-MIR	adjRI	0.497	0.497	0.497	0.497	0.497	0.498	0.497	0.497	0.497	0.497	0.501	0.497	0.497	0.497	0.497	**1.000**	0.498	0.497	0.497	0.497	**1.000**	**1.000**	0.500	0.499	0.497
	normMI	0.837	0.837	**0.837**	**0.837**	**0.837**	0.837	0.837	0.837	0.837	**0.837**	0.840	0.837	0.837	0.837	0.837	**1.443**	0.837	0.837	0.837	0.837	**1.443**	**1.443**	0.839	0.838	0.837
	purity	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.697	0.680	0.680	0.680	0.680	**1.000**	0.687	0.680	0.680	0.680	**1.000**	**1.000**	0.693	0.690	0.680
Gonen's MKK	adjRI	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497
	normMI	0.837	0.837	**0.837**	**0.837**	**0.837**	0.837	0.837	0.837	0.837	**0.837**	0.837	0.837	0.837	0.837	0.837	0.837	0.837	0.837	0.837	0.837	0.837	0.837	0.837	0.837	0.837
	purity	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680	0.680
Gonen's LMKK	adjRI	0.498	0.498	**0.498**	**0.498**	**0.498**	0.498	0.498	**0.498**	0.498	**0.498**	0.497	0.497	0.498	0.498	0.498	0.497	0.497	0.498	0.498	0.498	0.497	0.497	0.497	0.497	0.497
	normMI	0.837	0.837	**0.837**	**0.837**	**0.837**	0.837	0.837	0.837	0.837	**0.837**	0.837	0.837	0.837	0.837	0.837	0.837	0.837	0.837	0.837	0.837	0.837	0.837	0.837	0.837	0.837
	purity	0.683	0.683	**0.683**	**0.683**	**0.683**	0.683	0.683	**0.683**	0.683	**0.683**	0.680	0.680	0.683	0.683	0.683	0.680	0.680	0.683	0.683	0.683	0.680	0.680	0.680	0.680	0.680
Our Method	adjRI	0.497	**0.502**	0.497	0.497	0.497	**0.941**	**0.788**	**0.498**	**0.499**	0.497	**0.980**	**0.990**	**1.000**	**1.000**	**0.844**	**1.000**	**1.000**	**1.000**	0.980	**0.970**	**1.000**	**1.000**	**1.000**	**1.000**	0.970
	normMI	0.836	**0.841**	**0.837**	**0.837**	**0.837**	**1.325**	**1.127**	**0.838**	**0.838**	0.836	**1.400**	**1.418**	**1.443**	**1.443**	**1.203**	**1.443**	**1.443**	**1.443**	1.400	**1.384**	**1.443**	**1.443**	**1.443**	**1.443**	1.384
	purity	0.667	**0.700**	0.680	0.680	0.680	**0.980**	**0.920**	**0.683**	**0.690**	0.677	**0.993**	**0.997**	**1.000**	**1.000**	**0.943**	**1.000**	**1.000**	**1.000**	0.993	**0.990**	**1.000**	**1.000**	**1.000**	**1.000**	0.990

Clustering performance of the methods are evaluated by three widely-used metrics: Adjusted Rand Index (adjRI), Normalized Mutual Information (normMI), and purity. A higher value of the metrics indicates better clustering performance. Each column represents a simulated data set where the corresponding number indicates the number of the redundant variables (**N**$_{redun}$) added to the complete view (View 1) and correlation between each the redundant variables and the original variables. The bolded numbers are the maximum value for each the evalution measure within a simulation data set.

Table S5: **Evaluation of the Clutering Methods on Scenario B-Redun**

Scenario B-2																											
	cor	**0.45**					**0.72**					**0.90**					**0.97**					**1**				
	N$_{redun}$	**2**	**4**	**6**	**8**	**10**	**2**	**4**	**6**	**8**	**10**	**2**	**4**	**6**	**8**	**10**	**2**	**4**	**6**	**8**	**10**	**2**	**4**	**6**	**8**	**10**
Single Best	adjRI	0.487	**0.487**	**0.487**	**0.487**	**0.487**	0.487	0.487	**0.487**	**0.487**	**0.487**	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487
	normMI	0.811	0.811	**0.811**	**0.811**	**0.811**	0.811	0.811	**0.811**	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811
	purity	0.663	0.663	0.663	**0.663**	**0.663**	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663
Uniform Weight	adjRI	0.487	**0.487**	**0.487**	**0.487**	**0.487**	0.914	0.487	**0.487**	**0.487**	**0.487**	**0.990**	0.951	0.497	0.487	0.487	**1.000**	**1.000**	**1.000**	**0.990**	0.961	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
	normMI	0.812	**0.812**	**0.811**	**0.811**	**0.811**	1.288	0.811	**0.811**	0.811	0.811	**1.418**	1.355	0.837	0.811	0.811	**1.443**	**1.443**	**1.443**	**1.418**	1.359	**1.443**	**1.443**	**1.443**	**1.443**	**1.443**
	purity	0.670	0.670	0.663	**0.663**	**0.663**	0.970	0.670	**0.667**	0.663	0.663	**0.997**	0.983	0.680	0.667	0.667	**1.000**	**1.000**	**1.000**	**0.997**	0.987	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
MinMax MinC	adjRI	0.487	**0.487**	**0.487**	**0.487**	**0.487**	0.914	0.487	**0.487**	**0.487**	**0.487**	**0.990**	0.951	0.497	0.487	0.487	**1.000**	**1.000**	**1.000**	**0.990**	0.961	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
	normMI	0.812	**0.812**	**0.811**	**0.811**	**0.811**	1.288	0.811	**0.811**	0.811	0.811	**1.418**	1.355	0.837	0.811	0.811	**1.443**	**1.443**	**1.443**	**1.418**	1.359	**1.443**	**1.443**	**1.443**	**1.443**	**1.443**
	purity	0.670	0.670	0.663	**0.663**	**0.663**	0.970	0.670	**0.667**	0.663	0.663	**0.997**	0.983	0.680	0.667	0.667	**1.000**	**1.000**	**1.000**	**0.997**	0.987	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
Yu's OKKC	adjRI	0.487	**0.487**	**0.487**	**0.487**	**0.487**	0.487	0.487	**0.487**	**0.487**	**0.487**	0.487	0.487	0.487	0.487	0.487	0.869	0.487	0.487	0.487	0.487	**1.000**	0.519	0.487	0.487	0.487
	normMI	0.811	0.811	**0.811**	**0.811**	**0.811**	0.811	0.811	**0.811**	0.811	0.811	0.811	0.811	0.811	0.811	0.811	1.221	0.811	0.811	0.811	0.811	**1.443**	0.855	0.811	0.811	0.811
	purity	0.663	0.663	0.663	**0.663**	**0.663**	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.953	0.663	0.663	0.663	0.663	**1.000**	0.737	0.663	0.663	0.663
Liu's MKK-MIR	adjRI	0.487	**0.487**	**0.487**	**0.487**	**0.487**	0.487	0.487	**0.487**	**0.487**	**0.487**	0.487	0.487	0.487	0.487	0.487	**1.000**	0.487	0.487	0.487	0.487	**1.000**	**1.000**	**1.000**	0.487	0.487
	normMI	0.811	0.811	**0.811**	**0.811**	**0.811**	0.811	0.811	**0.811**	0.811	0.811	0.812	0.811	0.811	0.811	0.811	**1.443**	0.812	0.811	0.811	0.811	**1.443**	**1.443**	**1.443**	0.812	0.812
	purity	0.663	0.663	0.663	**0.663**	**0.663**	0.663	0.663	0.663	0.663	0.663	0.670	0.663	0.663	0.663	0.663	**1.000**	0.670	0.663	0.663	0.663	**1.000**	**1.000**	**1.000**	0.670	0.670
Gonen's MKK	adjRI	0.487	**0.487**	**0.487**	**0.487**	**0.487**	0.487	0.487	**0.487**	**0.487**	**0.487**	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487
	normMI	0.811	0.811	**0.811**	**0.811**	**0.811**	0.811	0.811	**0.811**	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811
	purity	0.663	0.663	0.663	**0.663**	**0.663**	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663
Gonen's LMKK	adjRI	0.487	**0.487**	**0.487**	**0.487**	**0.487**	0.487	0.487	**0.487**	**0.487**	**0.487**	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.614	0.618	0.487	0.487	0.487
	normMI	0.811	0.811	**0.811**	**0.811**	**0.811**	0.811	0.811	**0.811**	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.812	0.812	0.811	0.811	0.811	0.964	0.965	0.811	0.811	0.811
	purity	0.663	0.663	0.663	**0.663**	**0.663**	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.670	0.670	0.663	0.663	0.663	0.860	0.863	0.663	0.663	0.663
Our Method	adjRI	**0.632**	**0.487**	**0.487**	**0.487**	**0.487**	**0.961**	**0.590**	**0.487**	**0.487**	**0.487**	**0.990**	**0.980**	**0.990**	**0.980**	**0.914**	**1.000**	**1.000**	**1.000**	**0.990**	**0.980**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
	normMI	**1.009**	**0.812**	**0.811**	**0.811**	**0.811**	**1.369**	**0.921**	**0.811**	**0.812**	**0.812**	**1.418**	**1.394**	**1.418**	**1.400**	**1.308**	**1.443**	**1.443**	**1.443**	**1.418**	**1.394**	**1.443**	**1.443**	**1.443**	**1.443**	**1.443**
	purity	**0.840**	**0.673**	**0.667**	**0.663**	**0.663**	**0.987**	**0.810**	**0.667**	**0.670**	**0.673**	**0.997**	**0.993**	**0.997**	**0.993**	**0.970**	**1.000**	**1.000**	**1.000**	**0.997**	**0.993**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**

Clustering performance of the methods are evaluated by three widely-used metrics: Adjusted Rand Index (adjRI), Normalized Mutual Information (normMI), and purity. A higher value of the metrics indicates better clustering performance. Each column represents a simulated data set where the number indicates the number of the redundant variables (**N**$_{redun}$) added to the first partial view (View 1) and correlation between each the redundant variables and the original variables. The bolded numbers are the maximum value for each the evaluation measure within a simulation data set.

Table S6: **Evaluation of the Clutering Methods on Scenario C-Redun**

Scenario C-2																											
	cor	**0.45**					**0.72**					**0.90**					**0.97**					**1**				
	N$_{redun}$	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
Single Best	adjRI	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487
	normMI	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811
	purity	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663
Uniform Weight	adjRI	0.488	0.487	0.487	0.487	0.487	0.914	0.487	0.487	0.487	0.487	**0.990**	0.970	0.494	0.487	0.487	**1.000**	**1.000**	**1.000**	**0.990**	0.970	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
	normMI	0.812	0.811	0.811	0.811	0.811	1.295	0.812	0.812	0.812	0.811	**1.418**	1.384	0.818	0.812	0.812	**1.443**	**1.443**	**1.443**	**1.418**	1.375	**1.443**	**1.443**	**1.443**	**1.443**	**1.443**
	purity	0.673	0.667	0.670	0.667	0.667	0.970	0.677	0.667	0.667	0.667	**0.997**	0.990	0.703	0.673	0.677	**1.000**	**1.000**	**1.000**	**0.997**	0.990	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
MinMax MinC	adjRI	0.488	0.487	0.487	0.487	0.487	0.914	0.487	0.487	0.487	0.487	**0.990**	0.970	0.494	0.487	0.487	**1.000**	**1.000**	**1.000**	**0.990**	0.970	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
	normMI	0.812	0.811	0.811	0.811	0.811	1.295	0.812	0.812	0.812	0.811	**1.418**	1.384	0.818	0.812	0.812	**1.443**	**1.443**	**1.443**	**1.418**	1.375	**1.443**	**1.443**	**1.443**	**1.443**	**1.443**
	purity	0.673	0.667	0.670	0.667	0.667	0.970	0.677	0.667	0.667	0.667	**0.997**	0.990	0.703	0.673	0.677	**1.000**	**1.000**	**1.000**	**0.997**	0.990	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
Yu's OKKC	adjRI	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.869	0.487	0.487	0.487	0.487	**1.000**	0.519	0.487	0.487	0.487
	normMI	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	1.221	0.811	0.811	0.811	0.811	**1.443**	0.855	0.811	0.811	0.811
	purity	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.953	0.663	0.663	0.663	0.663	**1.000**	0.737	0.663	0.663	0.663
Liu's MKK-MIR	adjRI	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	**1.000**	0.487	0.487	0.487	0.487	**1.000**	**1.000**	**1.000**	0.487	0.487
	normMI	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.812	0.812	0.811	0.811	0.811	**1.443**	0.812	0.811	0.811	0.811	**1.443**	**1.443**	**1.443**	0.812	0.812
	purity	0.663	0.663	0.663	0.663	0.663	0.667	0.663	0.663	0.663	0.663	0.670	0.670	0.667	0.663	0.663	**1.000**	0.670	0.663	0.663	0.663	**1.000**	**1.000**	**1.000**	0.670	0.670
Gonen's MKK	adjRI	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487	0.487
	normMI	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811	0.811
	purity	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663	0.663
Gonen's LMKK	adjRI	0.490	0.490	0.490	0.490	0.490	0.490	0.490	0.490	0.490	0.490	0.490	0.490	0.490	0.490	0.490	0.487	0.490	0.490	0.490	0.490	0.614	0.652	0.490	0.490	0.490
	normMI	0.814	0.814	0.814	0.814	0.814	0.814	0.814	0.814	0.814	0.814	0.814	0.814	0.814	0.814	**0.814**	0.811	0.814	0.814	0.814	0.814	0.964	0.984	0.814	0.814	0.814
	purity	0.683	**0.683**	**0.683**	**0.683**	**0.683**	0.683	0.683	**0.683**	**0.683**	**0.683**	0.683	0.683	0.683	0.683	0.683	0.663	0.683	0.683	0.683	0.683	0.860	0.870	0.683	0.683	0.683
Our Method	adjRI	**0.592**	**0.498**	**0.498**	**0.498**	**0.498**	**0.941**	**0.656**	**0.498**	**0.498**	**0.498**	**0.990**	**0.990**	**0.980**	**0.827**	**0.499**	**1.000**	**1.000**	0.990	**0.990**	**0.980**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
	normMI	**0.913**	**0.837**	**0.837**	**0.837**	**0.837**	**1.343**	**0.988**	**0.838**	**0.837**	**0.837**	**1.418**	**1.418**	**1.400**	**1.168**	0.801	**1.443**	**1.443**	1.418	**1.418**	**1.400**	**1.443**	**1.443**	**1.443**	**1.443**	**1.443**
	purity	**0.813**	0.680	0.680	0.680	0.680	**0.980**	**0.853**	**0.683**	0.673	0.680	**0.997**	**0.997**	**0.993**	**0.937**	**0.743**	**1.000**	**1.000**	0.997	**0.997**	**0.993**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**

Clustering performance of the methods are evaluated by three widely-used metrics: Adjusted Rand Index (adjRI), Normalized Mutual Information (normMI), and purity. A higher value of the metrics indicates better clustering performance. Each column represents a simulated data set where the number indicates the number of the redundant variables (**N**$_{redun}$) added to the complete view and correlation between each the redundant variables and the original variables. The bolded numbers are the maximum value for each the evalution measure within a simulation data set.

Table S7: **List of BRCA/GBM related KEGG pathway**

Group	Size	Pathway
Breast caner	8	Estrogen signaling pathway; PI3K-Akt signaling pathway; Notch signaling pathway; Wnt signaling pathway; Homologous recombination; MAPK signaling pathway; p53 signaling pathway
Cell cycle Glioma	7	MAPK signaling pathway; p53 signaling pathway; Cell cycle; Cytokine-cytokine receptor interaction; ErbB signaling pathway; Calcium signaling pathway; mTOR signaling pathway
Pathways in cancer	20	Estrogen signaling pathway; PI3K-Akt signaling pathway; Notch signaling pathway; Wnt signaling pathway; MAPK signaling pathway; p53 signaling pathway; Cell cycle; Adherens junction; ECM-receptor interaction; Focal adhesion; cAMP signaling pathway; Jak-STAT signaling pathway; Hedgehog signaling pathway; HIF-1 signaling pathway; VEGF signaling pathway; Apoptosis; TGF-beta signaling pathway; Cytokine-cytokine receptor interaction; Calcium signaling pathway; mTOR signaling pathway
Central carbon metabolism in cancer	10	MAPK signaling pathway; PI3K-Akt signaling pathway; mTOR signaling pathway; HIF-1 signaling pathway; Alanine, aspartate and glutamate metabolism; Citrate cycle (TCA cycle); Fatty acid biosynthesis; Glycolysis / Gluconeogenesis; Glycine, serine and threonine metabolism; Oxidative phosphorylation

The BRCA/GBM related biological pathways are provided by KEGG Pathway Database (`https://www.kegg.jp`), which defined independently from our data analysis. The list of BRCA related pathways is consist of pathways from Breast caner, Pathways in cancer, and Central carbon metabolism in caner. The list of GBM related pathways is consist of pathways from Glioma, Pathways in cancer, and Central carbon metabolism in cancer.

Table S8: **BRCA related KEGG pathways identified by our method**

P-value	Cluster	Enriched Pathway
0.080	3	Cell cycle
0.098	4	Alanine, aspartate and glutamate metabolism
0.037	5	Focal adhesion
0.043	5	Cytokine-cytokine receptor interaction
0.049	5	ECM-receptor interaction

BRCA related KEGG pathways identified by our method ($pvalue < 0.1$) for each cluster and p-values from the pathway enrichment analysis are reported. P-values were adjusted to control the false discovery rate using the Benjamini-Hochberg procedure Benjamini and Hochberg (1995).

Table S9: **GBM related KEGG pathways identified by our method**

P-value	Cluster	Enriched Pathway
0.000	1	Focal adhesion
0.001	1	ECM-receptor interaction
0.022	1	MAPK signaling pathway
0.056	1	ErbB signaling pathway
0.057	1	Calcium signaling pathway
0.095	1	Adherens junction
0.007	2	Focal adhesion
0.034	2	ECM-receptor interaction
0.090	3	Calcium signaling pathway
0.030	4	Calcium signaling pathway
0.072	4	MAPK signaling pathway

BRCA related KEGG pathways identified by our method ($pvalue < 0.1$) for each cluster and p-values from the pathway enrichment analysis are reported. P-values were adjusted to control the false discovery rate using the Benjamini-Hochberg procedure Benjamini and Hochberg (1995).

References

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *JRSS B*, pages 289–300.

Fan, K. (1949). On a theorem of weyl concerning eigenvalues of linear transformations i. *Proc Natl Acad Sci*, **35**(11).

Gönen, M. and Margolin, A. A. (2014). Localized data fusion for kernel k-means clustering with application to cancer biology. *NeurIPS*.

Hastie, T., Tibshirani, R., Sherlock, G., Eisen, M., Brown, P., and Botstein, D. (1999). Imputing missing data for gene expression arrays.

Kloft, M., Brefeld, U., Sonnenburg, S., and Zien, A. (2011). l_p-norm multiple kernel learning. *J Mach Learn Res*, **12**(Mar), 953–997.

Liu, X., Dou, Y., Yin, J., Wang, L., and Zhu, E. (2016). Multiple kernel k-means clustering with matrix-induced regularization. *AAAI*, pages 1888–1894.

Ong, C. and Zien, A. (2008). An automated combination of kernels for predicting protein subcellular localization. *Algorithms in Bioinformatics*, pages 186–197.

Schölkopf, B., Smola, A., and Müller, K.-R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, **10**(5), 1299–1319.

Yu, S., Tranchevent, L., Liu, X., Glanzel, W., Suykens, J. A., De Moor, B., and Moreau, Y. (2012). Optimized data fusion for kernel k-means clustering. *IEEE Trans Pattern Anal Mach Intell*, **34**(5), 1031–1039.