Supporting Information for "Integrative Factorization of Bidimensionally Linked Matrices", by Jun Young Park and Eric F. Lock

This document provides supporting information for "Integrative Factorization of Bidimensionally Linked Matrices". Appendix A gives the details of the ISSVT algorithm, Appendix B gives proofs for novel results in the main document, Appendix C gives additional simulation results for the vertical integration context when data are generated without noise, Appendix D provides residual diagnostics for the TCGA data analysis. Appendix E provides additional plots showing subtype distinctions under alternative methods for the TCGA analysis.

A Proposed Algorithm (ISSVT)

The ISSVT algorithm, proceeds as follows proceeds as follows:

1. Initialize $\widehat{\mathbf{\Theta}} = \{\widehat{\mathbf{G}}_{00}, \widehat{\mathbf{R}}_{i0}, \widehat{\mathbf{C}}_{0j}, \widehat{\mathbf{I}}_{ij} \mid i = 1, \ldots, p, j = 1, \ldots, q\}$.

2. For $i = 1, \ldots, p$ and $j = 1, \ldots, q$, apply Proposition 1 to obtain a closed form solution of the following:

(a) $\widehat{\mathbf{G}}_{00}^{(new)} = \underset{\mathbf{G}_{00}:r(\mathbf{G}_{00})\leq r_{00}}{\arg\min} \frac{1}{2}||\mathbf{X}_{00} - \widehat{\mathbf{R}}_{00} - \widehat{\mathbf{C}}_{00} - \widehat{\mathbf{I}}_{00} - \mathbf{G}_{00}||_F^2 + \lambda_{00}||\mathbf{G}_{00}||_*$

(b) $\widehat{\mathbf{R}}_{i0}^{(new)} = \underset{\mathbf{R}_{i0}:r(\mathbf{R}_{i0})\leq r_{i0}}{\arg\min} \frac{1}{2}||\mathbf{X}_{i0} - \widehat{\mathbf{G}}_{i0}^{(new)} - \widehat{\mathbf{C}}_{i0} - \widehat{\mathbf{I}}_{i0} - \mathbf{R}_{i0}||_F^2 + \lambda_{i0}||\mathbf{R}_{i0}||_*$

(c) $\widehat{\mathbf{C}}_{0j}^{(new)} = \underset{\mathbf{C}_{0j}:r(\mathbf{C}_{0j})\leq r_{0j}}{\arg\min} \frac{1}{2}||\mathbf{X}_{0j} - \widehat{\mathbf{G}}_{0j}^{(new)} - \widehat{\mathbf{R}}_{0j}^{(new)} - \widehat{\mathbf{I}}_{0j} - \mathbf{C}_{0j}||_F^2 + \lambda_{0j}||\mathbf{C}_{0j}||_*$

(d) $\widehat{\mathbf{I}}_{ij}^{(new)} = \underset{\mathbf{I}_{ij}:r(\mathbf{I}_{ij})\leq r_{ij}}{\arg\min} \frac{1}{2}||\mathbf{X}_{ij} - \widehat{\mathbf{G}}_{ij}^{(new)} - \widehat{\mathbf{R}}_{ij}^{(new)} - \widehat{\mathbf{C}}_{ij}^{(new)} - \mathbf{I}_{ij}||_F^2 + \lambda_{ij}||\mathbf{I}_{ij}||_*$

3. The algorithm converges if $f_2(\widehat{\mathbf{\Theta}}) - f_2(\widehat{\mathbf{\Theta}}^{(new)}) < \epsilon$. If it does not converge, replace $\widehat{\mathbf{\Theta}}^{(new)}$ with $\widehat{\mathbf{\Theta}}$ and repeat Step 2.

The algorithm iteratively minimizes the objective f_2 (6) over blocks $\mathbf{G}_{00}$, $\{\mathbf{R}_{i0} \mid i = 1, \ldots, p\}$, $\{\mathbf{C}_{0j} \mid j = 1, \ldots, q\}$, and $\{\mathbf{I}_{ij} \mid i = 1, \ldots, p, j = 1, \ldots, q\}$. By Proposition 1, this is equivalent to a blockwise coordinate descent algorithm for f_1 (4), with corresponding update blocks $\{\mathbf{U}_{i0}^{(G)}, \mathbf{V}_{0j}^{(G)} \mid i = 1, \ldots, p, j = 1, \ldots, q\}$, $\{\mathbf{U}_{ij}^{(C)}, \mathbf{V}_{0j}^{(C)} \mid i = 1, \ldots, p, j = 1, \ldots, q\}$, $\{\mathbf{U}_{i0}^{(R)}, \mathbf{V}_{ij}^{(R)} \mid i = 1, \ldots, p, j = 1, \ldots, q\}$, $\{\mathbf{U}_{ij}^{(R)}, \mathbf{V}_{ij}^{(I)} \mid i = 1, \ldots, p, j = 1, \ldots, q\}$.

B Proofs

Theorem 1. *Let*

$$\widehat{\mathbf{\Theta}}_1 = \{\widehat{\mathbf{U}}_{i0}^{(G)}\widehat{\mathbf{V}}_{0j}^{(G)T}, \widehat{\mathbf{U}}_{i0}^{(R)}\widehat{\mathbf{V}}_{ij}^{(R)T}, \widehat{\mathbf{U}}_{ij}^{(C)}\widehat{\mathbf{V}}_{0j}^{(\mathbf{C})T}, \widehat{\mathbf{U}}_{ij}^{(I)}\widehat{\mathbf{V}}_{ij}^{(I)T} \mid i = 1 \ldots p, j = 1, \ldots q\}$$

minimize (4). Then,

$$\widehat{\mathbf{\Theta}}_2 = \{\widehat{\mathbf{G}}_{ij}, \widehat{\mathbf{R}}_{ij}, \widehat{\mathbf{C}}_{ij}, \widehat{\mathbf{I}}_{ij} \mid i = 1 \ldots p, j = 1, \ldots q\}$$

minimizes (6), where $\widehat{\mathbf{G}}_{ij} = \widehat{\mathbf{U}}_{i0}^{(G)}\widehat{\mathbf{V}}_{0j}^{(G)T}, \widehat{\mathbf{R}}_{ij} = \widehat{\mathbf{U}}_{i0}^{(R)}\widehat{\mathbf{V}}_{ij}^{(R)T}, \widehat{\mathbf{C}}_{ij} = \widehat{\mathbf{U}}_{ij}^{(C)}\widehat{\mathbf{V}}_{0j}^{(C)T}$, *and* $\widehat{\mathbf{I}}_{ij} = \widehat{\mathbf{U}}_{ij}^{(I)}\widehat{\mathbf{V}}_{ij}^{(I)T}$.

Proof. Because $\widehat{\boldsymbol{\Theta}}_1$ minimizes f_1,

$$||\widehat{\mathbf{U}}_{ij}^{(I)}||_F^2 + ||\widehat{\mathbf{V}}_{ij}^{(I)}||_F^2 = \min_{\substack{\mathbf{U}\mathbf{V}^T=\widehat{\mathbf{I}}_{ij} \\ \mathbf{U}:m_1\times\min(m_1,n_1) \\ \mathbf{V}:n_1\times\min(m_1,n_1)}} (||\mathbf{U}||_F^2 + ||\mathbf{V}||_F^2).$$

It follows from Lemma 1 that

$$||\widehat{\mathbf{U}}_{ij}^{(I)}||_F^2 + ||\widehat{\mathbf{V}}_{ij}^{(I)}||_F^2 = 2||\widehat{\mathbf{I}}_{ij}||_* \text{ for all } i > 0, j > 0.$$

Analogous arguments show that

$$\begin{aligned}
&||\widehat{\mathbf{U}}_{i0}^{(R)}||_F^2 + ||\widehat{\mathbf{V}}_{i0}^{(R)}||_F^2 = 2||\widehat{\mathbf{R}}_{i0}||_* \text{ for all } i > 0 \\
&||\widehat{\mathbf{U}}_{0j}^{(C)}||_F^2 + ||\widehat{\mathbf{V}}_{0j}^{(C)}||_F^2 = 2||\widehat{\mathbf{C}}_{0j}||_* \text{ for all } j > 0, \text{ and} \\
&||\widehat{\mathbf{U}}_{00}^{(G)}||_F^2 + ||\widehat{\mathbf{V}}_{00}^{(G)}||_F^2 = 2||\widehat{\mathbf{G}}_{00}||_*.
\end{aligned}$$

Thus, $f_1(\widehat{\boldsymbol{\Theta}}_1) = 2f_2(\widehat{\boldsymbol{\Theta}}_2)$.

Consider an alternative estimate $\widetilde{\boldsymbol{\Theta}}_2 = \{\widetilde{\mathbf{G}}_{ij}, \widetilde{\mathbf{R}}_{ij}, \widetilde{\mathbf{C}}_{ij}, \widetilde{\mathbf{I}}_{ij} \mid i = 1 \dots p, j = 1, \dots q\}$. By Lemma 1, $2f_2(\widetilde{\boldsymbol{\Theta}}_2) = f_1(\widetilde{\boldsymbol{\Theta}}_1)$ for some

$$\widetilde{\boldsymbol{\Theta}}_1 = \{\widetilde{\mathbf{U}}_{i0}^{(G)}\widetilde{\mathbf{V}}_{0j}^{(G)T}, \widetilde{\mathbf{U}}_{i0}^{(R)}\widetilde{\mathbf{V}}_{ij}^{(R)T}, \widetilde{\mathbf{U}}_{ij}^{(C)}\widetilde{\mathbf{V}}_{0j}^{(C)T}, \widetilde{\mathbf{U}}_{ij}^{(I)}\widetilde{\mathbf{V}}_{ij}^{(I)T} \mid i = 1 \dots p, j = 1, \dots q\}$$

where $\widetilde{\mathbf{G}}_{ij} = \widetilde{\mathbf{U}}_{i0}^{(G)}\widetilde{\mathbf{V}}_{0j}^{(G)T}, \widetilde{\mathbf{R}}_{ij} = \widetilde{\mathbf{U}}_{i0}^{(R)}\widetilde{\mathbf{V}}_{ij}^{(R)T}, \widetilde{\mathbf{C}}_{ij} = \widetilde{\mathbf{U}}_{ij}^{(C)}\widetilde{\mathbf{V}}_{0j}^{(C)T}$, and $\widetilde{\mathbf{I}}_{ij} = \widetilde{\mathbf{U}}_{ij}^{(I)}\widetilde{\mathbf{V}}_{ij}^{(I)T}$.

Thus, because $\widehat{\mathbf{\Theta}}_1$ minimizes f_1,

$$2f_2(\widetilde{\mathbf{\Theta}}_2) = f_1(\widetilde{\mathbf{\Theta}}_1) \geq f_1(\widehat{\mathbf{\Theta}}_1) = 2f_2(\widehat{\mathbf{\Theta}}_2),$$

and we conclude that $\widehat{\mathbf{\Theta}}_2$ minimizes f_2.

□

Theorem 2. *The objective $f_2(\cdot)$ in (6) is convex over its domain.*

Proof. Consider $\widetilde{\mathbf{\Theta}}^{(m)} = \{\mathbf{G}_{ij}^{(m)}, \mathbf{R}_{ij}^{(m)}, \mathbf{C}_{ij}^{(m)}, \mathbf{I}_{ij}^{(m)} \mid i = 1, \ldots, p, j = 1, \ldots q\}$ for $m = 1, 2$ and $\alpha \in [0, 1]$. It suffices to show

$$f_2\left(\alpha\widetilde{\mathbf{\Theta}}^{(1)} + (1-\alpha)\widetilde{\mathbf{\Theta}}^{(2)}\right) \leq \alpha f_2(\widetilde{\mathbf{\Theta}}^{(1)}) + (1-\alpha) f_2(\widetilde{\mathbf{\Theta}}^{(2)}), \tag{11}$$

where $\alpha\widetilde{\mathbf{\Theta}}^{(1)} + (1-\alpha)\widetilde{\mathbf{\Theta}}^{(2)}$ is given by

$$\{\alpha\mathbf{G}_{ij}^{(1)} + (1-\alpha)\mathbf{G}_{ij}^{(2)}, \alpha\mathbf{R}_{ij}^{(1)} + (1-\alpha)\mathbf{R}_{ij}^{(2)}, \alpha\mathbf{C}_{ij}^{(1)} + (1-\alpha)\mathbf{C}_{ij}^{(2)}, \alpha\mathbf{I}_{ij}^{(1)} + (1-\alpha)\mathbf{I}_{ij}^{(2)} \mid i = 1, \ldots, p, j = 1, \ldots q\}.$$

Decompose $f_2(\mathbf{\Theta}) = f_2^{\text{LS}}(\mathbf{\Theta}) + f_2^{\text{PEN}}(\mathbf{\Theta})$, where

$$f_2^{\text{LS}}(\mathbf{\Theta}) = \frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{q} ||\mathbf{X}_{ij} - \mathbf{G}_{ij} - \mathbf{R}_{ij} - \mathbf{C}_{ij} - \mathbf{I}_{ij}||_F^2, \text{ and}$$

$$f_2^{\text{PEN}}(\mathbf{\Theta}) = \lambda_{00}||\mathbf{G}_{00}||_* + \sum_{i=1}^{p}\lambda_{i0}||\mathbf{R}_{i0}||_* + \sum_{j=1}^{q}\lambda_{0j}||\mathbf{C}_{0j}||_* + \sum_{i=1}^{p}\sum_{j=1}^{q}\lambda_{ij}||\mathbf{I}_{ij}||_*.$$

By convexity of the least squares objective

$$||\mathbf{X}_{ij} - (\alpha\mathbf{A}_{ij}^{(1)} + (1-\alpha)\mathbf{A}_{ij}^{(2)})||_F^2 \leq \alpha||\mathbf{X}_{ij} - \mathbf{A}_{ij}^{(1)}||_F^2 + (1-\alpha)||\mathbf{X}_{ij} - \mathbf{A}_{ij}^{(2)}||_F^2 \tag{12}$$

for any $\mathbf{A}_{ij}^{(1)}$ and $\mathbf{A}_{ij}^{(2)}$. Applying (12) for each (i,j), where $\mathbf{A}_{ij}^{(m)} = \mathbf{G}_{ij}^{(m)} + \mathbf{R}_{ij}^{(m)} + \mathbf{C}_{ij}^{(m)} + \mathbf{I}_{ij}^{(m)}$ for $m = 1, 2$, gives

$$f_2^{\text{LS}}\left(\alpha\widetilde{\mathbf{\Theta}}^{(1)} + (1-\alpha)\widetilde{\mathbf{\Theta}}^{(2)}\right) \leq \alpha f_2^{\text{LS}}(\widetilde{\mathbf{\Theta}}^{(1)}) + (1-\alpha) f_2^{\text{LS}}(\widetilde{\mathbf{\Theta}}^{(2)}). \tag{13}$$

By convexity of the nuclear norm operator,

$$||\alpha\mathbf{A}^{(1)} + (1-\alpha)\mathbf{A}^{(2)}||_* \leq \alpha||\mathbf{A}^{(1)}||_* + (1-\alpha)||\mathbf{A}^{(2)}||_* \tag{14}$$

for any $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$. Applying (14) to each additive term in f_2^{PEN} gives

$$f_2^{\text{PEN}}\left(\alpha\widetilde{\mathbf{\Theta}}^{(1)} + (1-\alpha)\widetilde{\mathbf{\Theta}}^{(2)}\right) \leq \alpha f_2^{\text{PEN}}(\widetilde{\mathbf{\Theta}}^{(1)}) + (1-\alpha) f_2^{\text{PEN}}(\widetilde{\mathbf{\Theta}}^{(2)}). \tag{15}$$

Thus, (13) and (15) imply (11). □

Proposition 3. *The following conditions are necessary to allow for non-zero* $\widehat{\mathbf{G}}_{00}$, $\widehat{\mathbf{R}}_{i0}$, $\widehat{\mathbf{C}}_{0j}$, *and* $\widehat{\mathbf{I}}_{ij}$*:*

1. $max_j\lambda_{ij} < \lambda_{i0} < \sum_j \lambda_{ij}$ *for* $i = 1, \ldots, p$ *and* $max_i\lambda_{ij} < \lambda_{0j} < \sum_i \lambda_{ij}$ *for* $j = 1, \ldots, q$
2. $max_i\lambda_{i0} < \lambda_{00} < \sum_i \lambda_{i0}$ *and* $max_j\lambda_{0j} < \lambda_{00} < \sum_j \lambda_{0j}$.

Proof. Consider a violation of the left-hand inequality in condition 1.: $\lambda_{ij} \geq \lambda_{i0}$.

Define $\widehat{\boldsymbol{\Theta}}$ to be identical to $\widetilde{\boldsymbol{\Theta}}$ but with $\widehat{\mathbf{I}}_{ij} = \mathbf{0}$ and $\widehat{\mathbf{R}}_{ij} = \widetilde{\mathbf{R}}_{ij} + \widetilde{\mathbf{I}}_{ij}$. By convexity of the nuclear norm, $||\widehat{\mathbf{R}}_{i0}||_* \leq ||\widetilde{\mathbf{R}}_{i0}||_* + ||\widetilde{\mathbf{I}}_{ij}||_*$, and it follows that

$$\begin{aligned}
f_2(\widetilde{\boldsymbol{\Theta}}) - f_2(\widehat{\boldsymbol{\Theta}}) &= \lambda_{i0}||\widetilde{\mathbf{R}}_{i0}||_* + \lambda_{ij}||\widetilde{\mathbf{I}}_{ij}||_* - \lambda_{i0}||\widehat{\mathbf{R}}_{i0}||_* \\
&\geq \lambda_{i0}||\widetilde{\mathbf{R}}_{i0}||_* + \lambda_{ij}||\widetilde{\mathbf{I}}_{ij}||_* - \lambda_{i0}(||\widetilde{\mathbf{R}}_{i0}||_* + ||\widetilde{\mathbf{I}}_{ij}||_*) \\
&\geq \lambda_{i0}||\widetilde{\mathbf{R}}_{i0}||_* + \lambda_{ij}||\widetilde{\mathbf{I}}_{ij}||_* - \lambda_{i0}||\widetilde{\mathbf{R}}_{i0}||_* - \lambda_{ij}||\widetilde{\mathbf{I}}_{ij}||_* \\
&= 0.
\end{aligned}$$

Thus, regardless of the data $\mathbf{X}_{00}$, the objective $f_2(\boldsymbol{\Theta})$ is minimized with $\widehat{\mathbf{I}}_{ij} = \mathbf{0}_{m_i \times n_j}$. An analogous argument show that a violation of $\lambda_{ij} < \lambda_{0j}$ implies $\widehat{\mathbf{I}}_{ij} = \mathbf{0}_{m_i \times n_j}$ for some i, j. Moreover, analogous arguments show that a violation of $\max_i \lambda_{i0} < \lambda_{00}$ implies $\widehat{\mathbf{R}}_{i0} = \mathbf{0}_{m_i \times n_0}$ for some i, and that a violation of $\max_j \lambda_{0j} < \lambda_{00}$ implies $\widehat{\mathbf{C}}_{0j} = \mathbf{0}_{m_0 \times n_j}$ for some j.

Now, consider a violation of the right-hand inequality of condition 1.: $\lambda_{i0} \geq \sum_j \lambda_{ij}$. Define $\widehat{\boldsymbol{\Theta}}$ to be identical to $\widetilde{\boldsymbol{\Theta}}$ but with $\widehat{\mathbf{R}}_{i0} = \mathbf{0}_{m_i \times n_0}$ and $\widehat{\mathbf{I}}_{ij} = \widetilde{\mathbf{R}}_{ij} + \widetilde{\mathbf{I}}_{ij}$ for $j = 1, \ldots, q$. Then,

$$\begin{aligned}
f_2(\widetilde{\boldsymbol{\Theta}}) - f_2(\widehat{\boldsymbol{\Theta}}) &= \lambda_{i0}||\widetilde{\mathbf{R}}_{i0}||_* + \sum_j \lambda_{ij}||\widetilde{\mathbf{I}}_{ij}||_* - \sum_j \lambda_{ij}||\widehat{\mathbf{I}}_{ij}||_* \\
&\geq \lambda_{i0}||\widetilde{\mathbf{R}}_{i0}||_* + \sum_j \lambda_{ij}||\widetilde{\mathbf{I}}_{ij}||_* - \sum_j \lambda_{ij}(||\widetilde{\mathbf{I}}_{ij}||_* + ||\widetilde{\mathbf{R}}_{ij}||_*) \\
&= \lambda_{i0}||\widetilde{\mathbf{R}}_{i0}||_* - \sum_j \lambda_{ij}||\widetilde{\mathbf{R}}_{ij}||_* \\
&\geq \lambda_{i0}||\widetilde{\mathbf{R}}_{i0}||_* - \sum_j \lambda_{ij}||\widetilde{\mathbf{R}}_{i0}||_* \\
&\geq 0.
\end{aligned}$$

Thus, regardless of the data $\mathbf{X}_{00}$, the objective $f_2(\mathbf{\Theta})$ is minimized with $\widehat{\mathbf{R}}_{i0} = \mathbf{0}_{m_i \times n_0}$. An analogous argument show that a violation of $\sum_{0j} < \sum_i \lambda_{ij}$ implies $\widehat{\mathbf{C}}_{0j} = \mathbf{0}_{m_0 \times n_j}$ for some j. Moreover, analogous arguments show that a violation of $\lambda_{00} < \sum_i \lambda_{i0}$ or $\lambda_{00} < \sum_j \lambda_{0j}$ imply $\widehat{\mathbf{G}}_{00} = \mathbf{0}_{m_0 \times n_0}$. □ □

C Noiseless simulation

Here we generate joint and individual signals under the UNIFAC (vertically linked) model, and assess their recovery without residual error. We contrast the decomposition provided by the penalized objective $f_2(\cdot)$, with that obtained by alternative methods that enforce orthogonality of the estimated components.

We generate $\mathbf{X}_1 : d_1 \times n$ and $\mathbf{X}_2 : d_2 \times n$ as a sum of low-rank column-shared and individual structures that are *independent* but not necessarily *orthogonal*. That is,

$$\mathbf{X}_1 = \mathbf{C}_1 + \mathbf{I}_1$$

$$\mathbf{X}_2 = \mathbf{C}_2 + \mathbf{I}_2$$

where

$$\mathbf{C}_i = \mathbf{U}_i^{(C)}\mathbf{V}^T \text{ and } \mathbf{I}_i = \mathbf{U}_i^{(I)}\mathbf{V}_i^T \text{ for } i = 1, 2,$$

and the entries of $\mathbf{U}_1^{(C)} : d_1 \times r, \mathbf{U}_2^{(C)} : d_1 \times r, \mathbf{V} : n \times r, \mathbf{U}_1^{(I)} : d_1 \times r_1, \mathbf{V}_1 : n \times r_2, \mathbf{U}_2^{(I)} : d_2 \times r_2, \mathbf{V}_2 : n \times r_2$ are generated independently from a $\mathcal{N}(0, 1)$ distribution.

We fix $r = r_1 = r_2 = 10$ and generate 10 datasets under each of four scenarios with different row and column dimensions: (1) $d_1 = d_2 = n = 100$, (2) $d_1 = d_2 =$

$100, n = 500$, (3) $d_1 = d_2 = 500, n = 100$, (4) $d_1 = d_2 = n = 500$. For each generated dataset we apply (i) UNIFAC, (ii) JIVE, (iii) AJIVE, and (iv) SLIDE, where the correctly specified ranks are used for methods (ii–iv). For UNIFAC, the noise variance is set to a small value ($\sigma = 0.0001$) and the default penalties are used. In each case we compute the mean relative error in recovering underlying joint and individual structures:

$$\text{PredErr}(\widehat{\mathbf{C}}, \widehat{\mathbf{I}}) = \frac{1}{4}\left(\frac{||\mathbf{C}_1 - \widehat{\mathbf{C}}_1||_F^2}{||\mathbf{C}_1||_F^2} + \frac{||\mathbf{I}_1 - \widehat{\mathbf{I}}_1||_F^2}{||\mathbf{I}_1||_F^2} + \frac{||\mathbf{C}_2 - \widehat{\mathbf{C}}_2||_F^2}{||\mathbf{C}_2||_F^2} + \frac{||\mathbf{I}_2 - \widehat{\mathbf{I}}_2||_F^2}{||\mathbf{I}_2||_F^2}\right). \tag{16}$$

The results are summarized in Table 1. All methods decompose the underlying joint and individual signals with negligible error as the dimension of the sample size (n) and dimensions ($d = d_2 = d_2$) increase. UNIFAC recovers the joint and individual signals with comparable or substantially improved accuracy across the four scenarios, despite the use of the correct ranks for the other three methods. The error in recovery for methods (ii-iv) is due to the orthogonality constraints, which are necessary for identifiability of the decomposition without additional penalization. The independent joint and individual structures are not exactly orthogonal, but will approach orthogonality as $n \to \infty$, and thus performance improves for $n = 500$ vs $n = 100$. JIVE and AJIVE both assume orthogonality of the rows of the joint and individual structure, $\mathbf{C}$ and $\mathbf{I}$; thus, their performance is comparable across scenarios and differences are solely due to imprecision of the computational algorithm for JIVE. SLIDE additionally assumes orthogonality of the individual structures $\mathbf{I}_1$ and $\mathbf{I}_2$, which results in slightly less accurate recovery for lower n. Even under independence the underlying joint and individual signals

are not precisely orthogonal. The performance of UNIFAC for higher dimension d demonstrates the potential to recover the true joint and individual signals more accurately by relaxing orthogonality constraints.

Table 1: Mean relative recovery error for joint and individual signals, with standard error in parentheses, across different scenarios.

	$d = 100, n = 100$	$d = 500, n = 100$	$d = 100, n = 500$	$d = 500, n = 500$
UNIFAC	**0.058** (0.002)	**0.033** (0.001)	**0.027** (0.001)	**0.010** (0.001)
JIVE	**0.106** (0.003)	**0.110** (0.003)	**0.023** (0.001)	**0.022** (0.001)
AJIVE	**0.096** (0.003)	**0.100** (0.003)	**0.021** (0.001)	**0.021** (0.001)
SLIDE	**0.132** (0.005)	**0.130** (0.004)	**0.026** (0.001)	**0.024** (0.001)

We also compare the recovery of joint and individual signals as their ranks increase relative to the data dimensions. We consider the scenario with $d_1 = d_2 = 500, n = 100$, and generate 10 datasets for each of $r_1 = r_2 = r = \{1, 2, \ldots, 50\}$. For each dataset we estimate the decomposition via UNIFAC, AJIVE, or SLIDE; AJIVE is used instead of JIVE because in the noiseless scenario with given ranks they give the same underlying decomposition and AJIVE is more computationally efficient. The resulting mean relative errors (16) are shown in Figure 1. For ranks greater than $r = r_1 = r_2 = 33$, the sum of the ranks of column-shared and individual structures $(r + r_1 + r_2)$ is greater than the rank of the observed signal: $\text{rank}([\mathbf{X}_1^T, \mathbf{X}_2^T]) \leq 100$. Thus, for these cases a SLIDE decomposition with the given ranks does not exist because the condition of orthogonality between $\hat{J}$, $\hat{I}_1$ and $\hat{I}_2$ cannot be satisfied. The recovery errors for the AJIVE decomposition also increase sharply at this point, while the trend remains stable for UNIFAC. In general, UNIFAC provides better recovery of the generated joint and individual signal as the ranks increase. However, the recovery error does increase steadily as the ranks get larger. Moreover, if the sum of the ranks is greater than the

rank of the overall signal, this implies linear dependence among the underlying components (here, linear dependence among $\mathbf{V}$, $\mathbf{V}_1$, and $\mathbf{V}_2$), which complicates their interpretation.

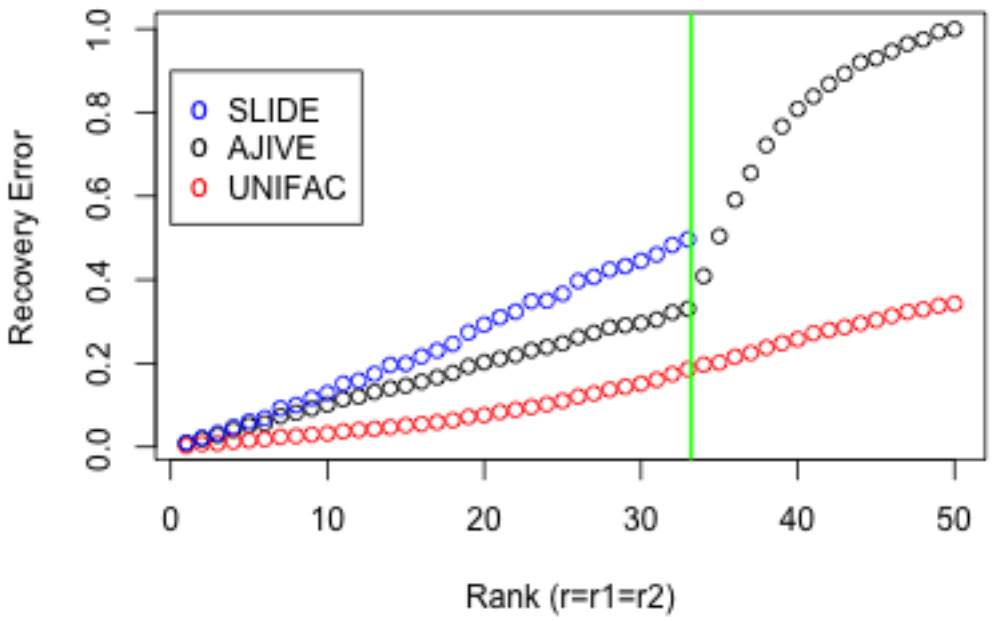

Figure 1: Mean relative recovery with data generated and estimated under the given ranks

D Data Analysis: Residual Diagnostics

Here we consider the distribution of residuals after the application of BIDIFAC for the TCGA application described in Section 4 of the main manuscript. That is,

we consider the residual matrices

$$\widehat{\mathbf{E}}_{ij} = \widehat{\mathbf{X}}_{ij} - \widehat{\mathbf{G}}_{ij} - \widehat{\mathbf{C}}_{ij} - \widehat{\mathbf{R}}_{ij} - \widehat{\mathbf{I}}_{ij}.$$

BIDIFAC is best motivated when the error terms $\mathbf{E}_{ij}$ are approximately Gaussian, as the objective identifies the mode of a Bayesian model with a Gaussian likelihood (Section 2.7) and the selection of the tuning parameters is based on the assumption of Gaussian error (Section 2.6). Figure 2 shows the distribution of residuals for each of the four datasets considered (tumor mRNA, tumor miRNA, normal mRNA, normal miRNA) overlayed with two Gaussian densities: one giving the theoretical distribution of residuals implied by the pre-hoc estimate of the noise variance for each dataset ($\hat{\sigma}_{ij}^{MAD}$), and another giving the empirical Gaussian fit resulting from the sample mean and standard deviation of the observed residuals. None of the residual histograms show strong departures from Gaussianity, and the variance estimate used to tune the model ($\hat{\sigma}_{ij}^{MAD}$) fits the observed residuals reasonably well in each case.

E Data Analysis: Scatterplots for competing methods

Here we show scatterplots of the shared gene-miRNA structure for the vertical integration approaches UNIFAC and JIVE, analogous to that in Figure 2 of the main manuscript. These plots also show prominent subtype distinctions, similar to that for BIDIFAC. However, the BIDIFAC analysis further demonstrates that these distinctions are primarily due to tumor-specific variability.

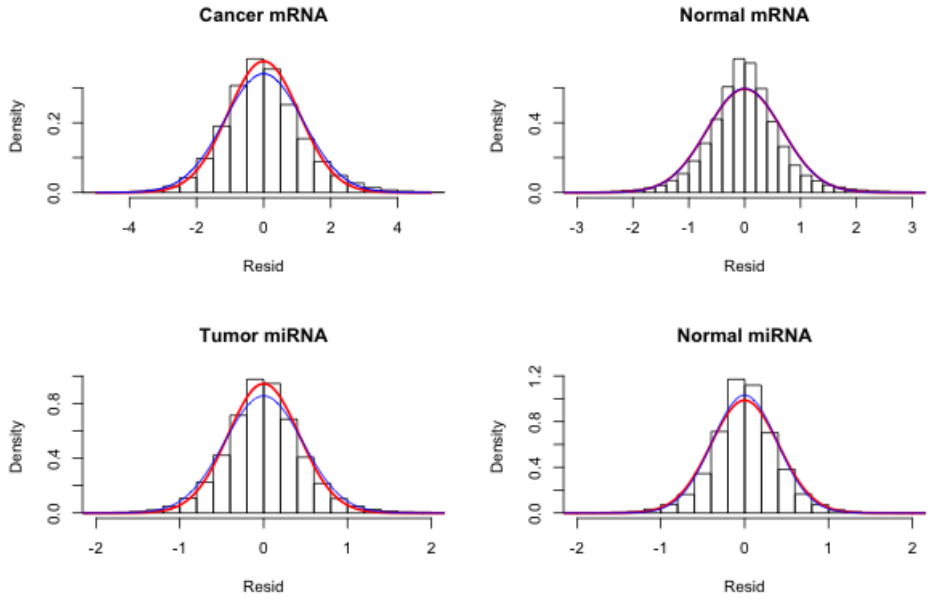

Figure 2: Distribution of residuals for the TCGA application. The Gaussian density given by the noise variance estimate $\hat{\sigma}_{ij}^{MAD}$ is shown in red, the density given by the emperical mean and standard deviation is shown in blue.

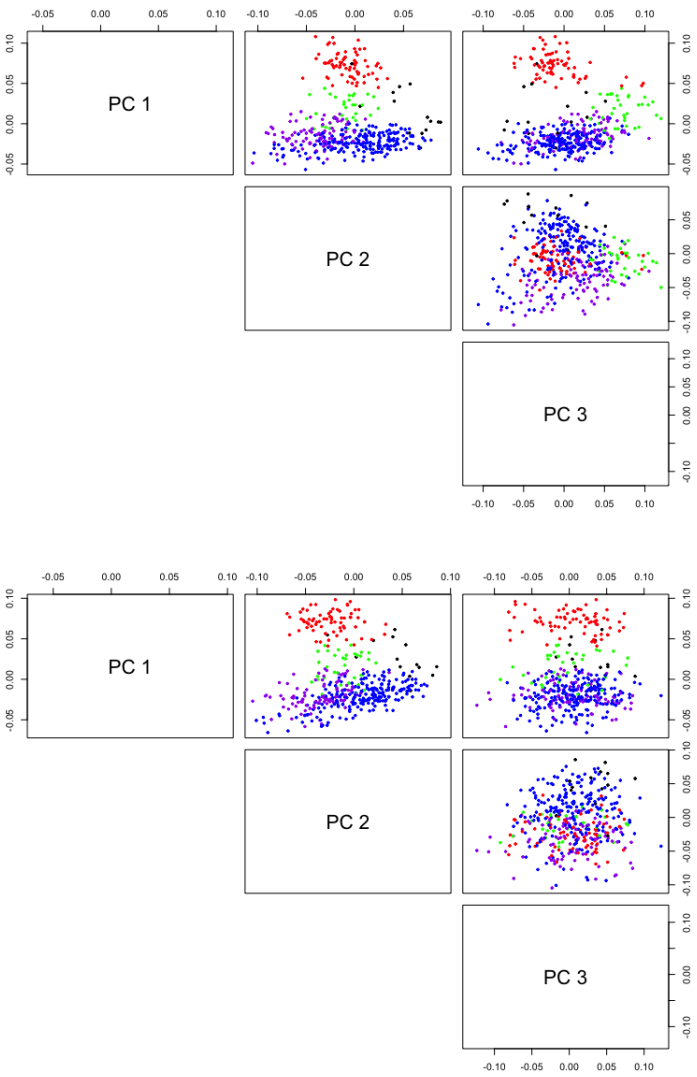

Figure 3: Scatterplots of top three principal components of the estimated column-shared structures: UNIFAC (top) and JIVE (bottom).