Appendix

A1 Transcriptome Data Preprocessing

We applied our methods to raw bulk RNA-Sequencing data of 44 matched primary breast and metastatic brain tumors from 22 patients (each patient gives two samples) [31,38], where six patients are from the Royal College of Surgeons (RCS) and sixteen patients from the University of Pittsburgh (Pitt). These data profiled the expression levels of approximately 60,000 transcripts. We removed the genes that are not expressed in any sample. We also considered only protein-coding genes in the present study. We conducted quantile normalization across samples using the geometric mean to remove possible artifacts from different experiment batches [1]. The top 2.5% and bottom 2.5% of expressions were clipped to further reduce noise. Finally, we transformed the resulting bulk gene expression values into the log space and mapped those for each gene to the interval $[0, 1]$ by a linear transformation.

A2 Mapping to Gene Modules and Cancer Pathways

The protein-coding gene expressions were mapped into both perturbed gene modules and cancer pathways, using the DAVID tool and external knowledge bases [18], as well as the cancer pathways in KEGG database [19]. This step compresses the high dimensional data and provides markers of cancer-related biological processes (Fig. 1 Step 1).

Gene Modules. Functionally similar genes are usually affected by a common set of somatic alterations [30] and therefore are co-expressed in the cells. These genes are believed to belong to the same "gene modules" [8,37]. Inspired by the idea of gene modules, we fed a subset of 3,000 most informative genes out of the approximately 20,000 genes that have the largest variances into the DAVID tool for functional annotation clustering using several databases [18]. DAVID maps each gene to one or more modules. We did not force the genes to be mapped into disjunct modules because a gene may be involved in several biological functions and therefore more than one gene module. We removed gene modules that were not enriched (fold enrichment < 1.0) and kept the remaining $m_1 = 109$ modules (and the corresponding annotated functions), where fold enrichment is defined as the EASE score of the current module to the geometric mean of EASE scores in all modules [17]. The gene module values of all the $n = 44$ samples were represented as a gene module matrix $\mathbf{B}_M \in \mathbb{R}^{m_1 \times n}$. The i-th gene module value in j-th sample, $\mathbf{B}_{M_{i,j}}$, was calculated by taking the sum of expressions of all the genes in the i-th module. Then $\mathbf{B}_M$ was rescaled row-wise by taking the z-scores across samples to compensate for the effect of variable module sizes.

Cancer Pathways. Although gene module representation is able to capture the variances across samples and reduce the redundancy of raw gene expressions, it has two disadvantages: First of all, lack of interpretability. Specifically, some annotations assigned by DAVID are not directly related to biological functions,

and the annotations of different modules may substantially overlap. Secondly, the key perturbed cancer pathways or functions may not be always the ones that vary most across samples. For example, genes in cancer-related KEGG pathways (hsa05200) [19] are not especially enriched in the top 3,000 genes with the largest expression variances. To make better use of prior knowledge on cancer-relevant pathways, we supplemented the generic DAVID pathway sets with a KEGG "cancer pathway" representation of samples $\mathbf{B}_P \in \mathbb{R}^{m_2 \times n}$, where the number of cancer pathways $m_2 = 24$. The cancer-related pathways in the KEGG database are cleaner and easier to explain, more orthogonal to each other, and contain critical signaling pathways to cancer development. We extracted the 23 cancer-related pathways from the following 3 KEGG pathway sets: *Pathways in cancer* (hsa05200), *Breast cancer* (hsa05224), and *Glioma* (hsa05214). An additional cancer pathway *RET pathway* was added, since it was found to be recurrently gained in the prior research [38]. See y-axis of Fig. 3d for the complete list of 24 cancer pathways. We considered all the ∼20,000 protein-coding genes other than top 3,000 genes. The following mapping of cancer pathways and transformation to z-scores were similar to that we did to map the gene modules.

Until this step, the raw gene expressions of n samples were transformed into the compressed gene module/pathway representation of samples $\mathbf{B} = \left[\mathbf{B}_M^\intercal, \mathbf{B}_P^\intercal\right]^\intercal \in \mathbb{R}^{m \times n}$, where $m = m_1 + m_2$. The gene module representation $\mathbf{B}_M$ serves for accurately deconvolving and unmixing the cell communities, while the pathway representation $\mathbf{B}_P$ serves as markers/probes and for interpretation purpose.

A3 Deconvolution of Bulk Data

A3.1 Non-convexity of Deconvolution Problem

Theorem 1. *The deconvolution problem Eq. (1-3) below is not convex:*

$$\min_{\boldsymbol{C},\boldsymbol{F}} \quad f(\boldsymbol{C},\boldsymbol{F}) = \|\boldsymbol{B} - \boldsymbol{C}\boldsymbol{F}\|_{\mathrm{Fr}}^2, \tag{A1}$$

$$s.t. \quad \boldsymbol{F}_{lj} \geq 0, \qquad l = 1,...,k,\ j = 1,...,n, \tag{A2}$$

$$\sum_{l=1}^{k} \boldsymbol{F}_{lj} = 1, \qquad j = 1,...,n. \tag{A3}$$

Proof. If the problem is convex, we should have: $\forall \lambda \in (0,1)$, and $\forall \mathbf{C}_x, \mathbf{C}_y, \mathbf{F}_x, \mathbf{F}_y$ in the feasible domain, the following inequality always holds:

$$\lambda f(\mathbf{C}_x, \mathbf{F}_x) + (1-\lambda) f(\mathbf{C}_y, \mathbf{F}_y) \geq f(\lambda \mathbf{C}_x + (1-\lambda)\mathbf{C}_y, \lambda \mathbf{F}_x + (1-\lambda)\mathbf{F}_y). \tag{A4}$$

However, for the following setting:

$$\mathbf{B} = \begin{bmatrix} -1.38 & 0.92 \\ 1.03 & -0.15 \end{bmatrix}, \tag{A5}$$

$$\mathbf{C}_x = \begin{bmatrix} -1.74 & 2.21 \\ 1.00 & -3.97 \end{bmatrix}, \qquad \mathbf{C}_y = \begin{bmatrix} 1.03 & -0.46 \\ -3.13 & 0.16 \end{bmatrix}, \tag{A6}$$

$$\mathbf{F}_x = \begin{bmatrix} 0.83 & 0.32 \\ 0.17 & 0.68 \end{bmatrix}, \qquad \mathbf{F}_y = \begin{bmatrix} 0.09 & 0.34 \\ 0.91 & 0.66 \end{bmatrix}, \tag{A7}$$

and $\lambda = 0.5$, we have

$$\lambda f(\mathbf{C}_x, \mathbf{F}_x) + (1-\lambda) f(\mathbf{C}_y, \mathbf{F}_y) = 4.86 < 11.74 = f(\lambda \mathbf{C}_x + (1-\lambda)\mathbf{C}_y, \lambda \mathbf{F}_x + (1-\lambda)\mathbf{F}_y). \tag{A8}$$

This is contradictory to Eq. (A4). □

A3.2 Architecture Specifications of NND

In the NND architecture, $|\mathbf{X}|$ applies element-wise absolute value, $\mathrm{cwn}(\mathbf{X})$ column-wisely normalizes $\mathbf{X}$, so that each column of the output sums up to 1. The two operations of Eq. (5) naturally rephrase and remove the two constraints in Eq. (2-3), and meanwhile fit the framework of neural networks. An alternative to the absolute value operation $|\mathbf{X}|$ might be rectified linear unit $\mathrm{ReLU}(\mathbf{X}) = \max(\mathbf{0}, \mathbf{X})$. However, this activation function is unstable and leads to inferior performance in our case, since $\mathbf{X}_{lj}$ will be fixed to zero once it becomes negative and will lose the chance to get updated in the following iterations. One may also want to replace the column-wise normalization $\mathrm{cwn}(\mathbf{X})$ with softmax operation $\mathrm{softmax}(\mathbf{X})$. However, the nonlinearity introduced by softmax actually changes the original optimization problem Eq. (1-3) and the fitted $\mathbf{F}$ is therefore not sparse.

A3.3 Hyperparameters of NND

We used an Adam optimizer with default momentum parameters and learning rate of 1×10^{-5} [20]. The mini-batch technique is not required since the data size in our application is small enough not to require it ($\mathbf{B} \in \mathbb{R}^{m \times n}$, $m = 133$, $n = 44$). The training was run until convergence, when the relative decrease of training loss is smaller than $\epsilon = 1 \times 10^{-10}$ every 20,000 iterations.

A3.4 Fitting Ability of NND

One might be suspicious whether the neural network fits precisely in practice, since it is based on a simple gradient descent optimization. To validate the fitting ability of NND, we plotted the PCA of original samples $\mathbf{B}$ and the fitted samples $\hat{\mathbf{B}} = \mathbf{CF}$ (Fig. A1). One can easily see that NND provides good model fits to the data.

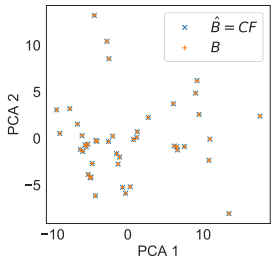

Fig. A1: PCA of pathway representation $\mathbf{B}$ and nnMF fitted $\hat{\mathbf{B}}$. Each dot represents the pathway values of a sample $\mathbf{B}_{\cdot j}$ or fitted $\hat{\mathbf{B}}_{\cdot j}$. The first two PCA dimensions of original data and fitted data are almost in the same positions, which indicates that NND is able to fit precisely in our application. The number of components is set to be $k = 5$ here.

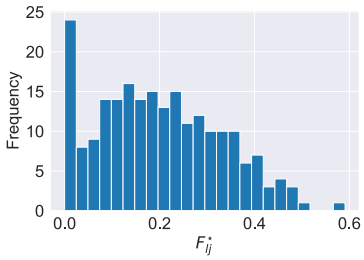

Fig. A2: Distribution of elements in fraction matrix $\mathbf{F}^{\star}$. Since each column of $\mathbf{F}$ is forced to sum up to be one, a Laplacian prior is applied to the elements of matrix $\mathbf{F}$. This leads to the sparsity of $\mathbf{F}^{\star}$: 24 out of its 220 elements ($k \times n = 5 \times 44$) are zeros (threshold set to 2.5×10^{-2}).

A3.5 Sparsity of NND Results

See Fig. A2 for distribution of fraction matrix in NND deconvolution results.

A3.6 Cross-validation of NND

In each fold of the CV, we used $\hat{\mathbf{B}} = \mathbf{CF}$ to only fit some randomly selected elements of $\mathbf{B}$, and then the test error was calculated using the other elements of $\mathbf{B}$. This was implemented by introducing two additional mask matrices $\mathbf{M}_{\text{train}}, \mathbf{M}_{\text{test}} \in \{0,1\}^{m \times n}$, which are in the same shape of $\mathbf{B}$, and $\mathbf{M}_{\text{train}} + \mathbf{M}_{\text{test}} = 1^{m \times n}$. During the training time, with the same constraints in Eq. (5), the optimization goal is:

$$\min_{\mathbf{C},\mathbf{F}_{\text{par}}} \|\mathbf{M}_{\text{train}} \odot (\mathbf{B} - \mathbf{CF})\|_{\text{Fr}}^2 \tag{A9}$$

where $\mathbf{X} \odot \mathbf{Y}$ is the Hadamard (element-wise) product. At the time of evaluation, given optimized $\mathbf{C}^\star$, $\mathbf{F}^\star_{\text{par}}$, and therefore optimized $\mathbf{F}^\star = \text{cwn}\left(\left|\mathbf{F}^\star_{\text{par}}\right|\right)$ for the optimization problem during training, the test error was calculated on the test set: $\|\mathbf{M}_{\text{test}} \odot (\mathbf{B} - \mathbf{C}^\star\mathbf{F}^\star)\|^2_{\text{Fr}}$. We used 20-fold cross-validation on the NND, so in each fold 95% positions of $\mathbf{M}_{\text{train}}$ and 5% positions of $\mathbf{M}_{\text{test}}$ were 1s.

A4 Derivation of Quadractic Programming, $\mathbf{P}(\mathcal{W})$, and $\mathbf{q}(\mathcal{W}, \mathbf{c})$

Recall Sec. 2.5, for the phylogeny $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, the Steiner nodes are indexed as $\mathcal{V}_S = \{1, 2, ..., k-2\}$ ($|\mathcal{V}_S| = k-2$), the extant nodes are indexed as $\mathcal{V}_C = \{k-1, k, ..., 2k-2\}$ ($|\mathcal{V}_C| = k$). The i-th pathway values of Steiner nodes are denoted as $\mathbf{x} = [x_1, x_2, ..., x_{k-2}]^\intercal \in \mathbb{R}^{k-2}$, and values of extant nodes as $\mathbf{y} = [y_{k-1}, y_k, ..., y_{2k-2}]^\intercal \in \mathbb{R}^k$. Since we consider each pathway dimension separately here, the subscript i for $\mathbf{x}$ and $\mathbf{y}$ is omitted for brevity. The weight of edge $(u, v) \in \mathcal{E}$ connecting nodes u and v is w_{uv} ($1 \le u < v \le 2k-2$). Denote $\mathcal{W} = \{w_{uv} \mid (u, v) \in \mathcal{E}\}$. The inference of the i-th element in the pathway vector of the Steiner nodes can be formulated as minimizing the elastic potential energy $U(\mathbf{x}, \mathbf{y}; \mathcal{W})$ shown below:

$$\min_{\mathbf{x}} \quad U(\mathbf{x}, \mathbf{y};\, \mathcal{W}) = \sum_{\substack{(u,v)\in\mathcal{E} \\ v \le k-2}} \frac{1}{2} w_{uv}(x_u - x_v)^2 + \sum_{\substack{(u,v)\in\mathcal{E} \\ v \ge k-1}} \frac{1}{2} w_{uv}(x_u - y_v)^2, \quad \text{(A10)}$$

Theorem 2. *Equation (A10) can be further rephrased as a quadratic programming problem:*

$$\min_{\boldsymbol{x}} \quad \frac{1}{2}\boldsymbol{x}^\intercal \boldsymbol{P}(\mathcal{W})\boldsymbol{x} + \boldsymbol{q}(\mathcal{W}, \boldsymbol{y})^\intercal \boldsymbol{x}, \quad \text{(A11)}$$

where $\boldsymbol{P}(\mathcal{W})$ is a function that takes as input edge weights $\mathcal{W}$ and outputs a matrix $\boldsymbol{P} \in \mathbb{R}^{(k-2)\times(k-2)}$, $\boldsymbol{q}(\mathcal{W}, \boldsymbol{y})$ is a function that takes as input edge weights $\mathcal{W}$ and vector $\boldsymbol{y}$ and outputs a vector $\boldsymbol{q} \in \mathbb{R}^{k-2}$.

Proof. Based on Eq. (A10), $U(\mathbf{x}, \mathbf{y}; \mathcal{W}) \ge 0$. Each term inside the first summation ($v \le k-2$) can be written as:

$$\frac{1}{2} w_{uv}(x_u - x_v)^2 = \frac{1}{2}\mathbf{x}^\intercal \mathbf{P}(w_{uv})\mathbf{x}, \quad \text{(A12)}$$

where

$$\mathbf{P}(w_{uv}) = \begin{array}{c} \\ u\text{-th row} \\ \\ v\text{-th row} \\ \end{array} \begin{array}{c} \begin{array}{ccccc} & u\text{-th col} & & v\text{-th col} & \end{array} \\ \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & w_{uv} & \mathbf{0} & -w_{uv} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -w_{uv} & \mathbf{0} & w_{uv} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \end{array}. \quad \text{(A13)}$$

Each term ($v \geq k-1$) inside the second summation can be rephrased as:

$$\frac{1}{2}w_{uv}(x_u - y_v)^2 = \frac{1}{2}\mathbf{x}^\intercal \mathbf{P}(w_{uv})\mathbf{x} + \mathbf{q}(w_{uv}, y_v)^\intercal \mathbf{x} + C(w_{uv}, y_v), \tag{A14}$$

where

$$\mathbf{P}(w_{uv}) = \begin{matrix} \\ u\text{-th row} \end{matrix} \overset{\begin{matrix} u\text{-th col} \end{matrix}}{\begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0\,0\,0} \\ \mathbf{0} & w_{uv} & \mathbf{0\,0\,0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0\,0\,0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0\,0\,0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0\,0\,0} \end{bmatrix}}, \quad \mathbf{q}(w_{uv}, y_v) = \begin{matrix} \\ u\text{-th row} \end{matrix} \begin{bmatrix} \mathbf{0} \\ -w_{uv}y_v \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \tag{A15}$$

and $C(w_{uv}, y_v) = \frac{1}{2}w_{uv}y_v^2$ is independent of $\mathbf{x}$. Therefore the optimization in Eq. (A10) can be calculated and written as below:

$$\min_{\mathbf{x}} \quad \sum_{\substack{(u,v)\in\mathcal{E} \\ v\leq k-2}} \frac{1}{2}\mathbf{x}^\intercal \mathbf{P}(w_{uv})\mathbf{x} + \sum_{\substack{(u,v)\in\mathcal{E} \\ v\geq k-1}} \left(\frac{1}{2}\mathbf{x}^\intercal \mathbf{P}(w_{uv})\mathbf{x} + \mathbf{q}(w_{uv}, y_v)^\intercal \mathbf{x}\right), \tag{A16}$$

$$\Leftrightarrow \min_{\mathbf{x}} \quad \frac{1}{2}\mathbf{x}^\intercal \left(\sum_{\substack{(u,v)\in\mathcal{E} \\ v\leq k-2}} \mathbf{P}(w_{uv}) + \sum_{\substack{(u,v)\in\mathcal{E} \\ v\geq k-1}} \mathbf{P}(w_{uv})\right)\mathbf{x} + \sum_{\substack{(u,v)\in\mathcal{E} \\ v\geq k-1}} \mathbf{q}(w_{uv}, y_v)^\intercal \mathbf{x}, \tag{A17}$$

$$\Leftrightarrow \min_{\mathbf{x}} \quad \frac{1}{2}\mathbf{x}^\intercal \mathbf{P}(\mathcal{W})\mathbf{x} + \mathbf{q}(\mathcal{W}, \mathbf{y})^\intercal \mathbf{x}. \square \tag{A18}$$

Remark 1. The optimal $\mathbf{x}^\star$ of the Eq. (A10), or the solution to the quadratic programming problem Eq. (A11) can be solved by setting the gradient to be $\mathbf{0}$:

$$\mathbf{P}(\mathcal{W})\mathbf{x}^\star + \mathbf{q}(\mathcal{W}, \mathbf{y}) = \mathbf{0}. \tag{A19}$$

Therefore,

$$\mathbf{x}^\star = -\mathbf{P}(\mathcal{W})^{-1}\mathbf{q}(\mathcal{W}, \mathbf{y}). \tag{A20}$$

Remark 2. Based on the proof, we can derive how to calculate the matrix $\mathbf{P}(\mathcal{W})$ and vector $\mathbf{q}(\mathcal{W}, \mathbf{y})$.

Initialize the matrix and vector with zeros:

$$\mathbf{P} \leftarrow 0^{(k-2)\times(k-2)}, \quad \mathbf{q} \leftarrow 0^{k-2}. \tag{A21}$$

For each edge $(u, v) \in \mathcal{E}$ with weight w_{uv}, there are two possibilities of nodes u and v: First, if both of them are Steiner nodes ($u \leq k-2$, $v \leq k-2$), we update $\mathbf{P}$ and keep $\mathbf{q}$ the same:

$$\mathbf{P}_{uu} \leftarrow \mathbf{P}_{uu} + w_{uv},\ \mathbf{P}_{vv} \leftarrow \mathbf{P}_{vv} + w_{uv},\ \mathbf{P}_{uv} \leftarrow \mathbf{P}_{uv} - w_{uv},\ \mathbf{P}_{vu} \leftarrow \mathbf{P}_{vu} - w_{uv}. \tag{A22}$$

Second, if u is Steiner node and v is an extant node ($u \leq k-2$, $v \geq k-1$), we update both $\mathbf{P}$ and $\mathbf{q}$:

$$\mathbf{P}_{uu} \leftarrow \mathbf{P}_{uu} + w_{uv}, \quad \mathbf{q}_u \leftarrow \mathbf{q}_u - y_v \cdot w_{uv}. \tag{A23}$$

We apply the same procedure to all dimension of pathways $i = 1, 2, ..., m$ to get the full pathway values for each Steiner node.

A5 Differentially Expressed Cancer Pathways

Table A1 provides a list of the identified differentially expressed cancer pathways.

Table A1: Differentially expressed cancer pathways between primary and metastatic samples (FDR<0.05).

Gain/Loss after metastasis	Differentially expressed pathways	FDR
Relative gain	cAMP signaling pathway	6.88e-03
Relative gain	ErbB signaling pathway	2.09e-02
Relative gain	Calcium signaling pathway	4.39e-02
Relative loss	Cytokine-cytokine receptor interaction	4.37e-06
Relative loss	Apoptosis	8.53e-04
Relative loss	JAK-STAT signaling pathway	8.53e-04
Relative loss	Wnt signaling pathway	3.97e-03
Relative loss	Hedgehog signaling pathway	4.50e-03
Relative loss	PI3K-Akt signaling pathway	1.35e-02
Relative loss	TGF-beta signaling pathway	4.56e-02
Relative loss	Notch signaling pathway	4.56e-02

A6 Portions of Cell Communities in BrM Patients

Fig. A3 shows the inferred cell community portions across the BrM samples. The figure displays, for each patient, the proportion of each community in the primary and the metastatic sample.

A7 Perturbed Cancer Pathways along Phylogenies

Table A2-A5 provide a full list of perturbed pathways across the phylogenies for Case 1, 2, 3, and 4 in Fig. 3e.

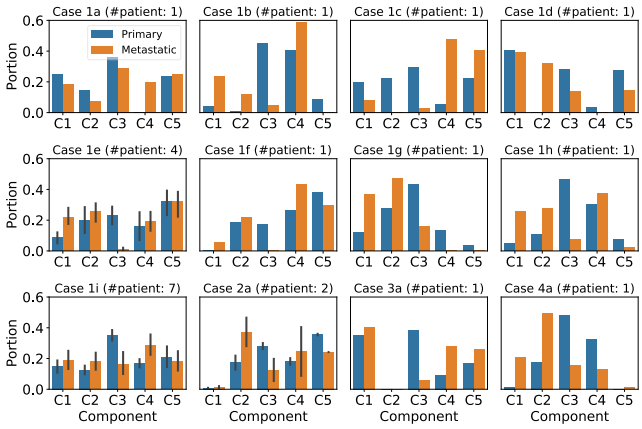

Fig. A3: Classification of BrM patients based on the consisted cell subcommunities in matched samples. There are 12 subcases of the 4 cases mentioned in Sec. 3.2. Specifically, there are 9 specific cases (Case 1a-i) in Case 1. Most patients (7) have all the five cell communities in both primary and metastatic samples (Case 1i). A few patients (4) have all communities in metastasis samples and all clones but community $C3|P$ in primary samples. The element $\mathbf{F}_{lj}$ is taken as 0 when it is smaller than a threshold 2.5×10^{-2}, and therefore the l-th community is missing in the j-th sample.

Table A2: Perturbed pathways during the evolution of cell communities in primary and metastatic tumors (Fig. 3e Case 1). The top five perturbed pathways whose gain or loss greater than 1.0 along each edge of phylogeny are shown. Clinically actionable perturbed cancer pathways during metastasis are shown in boldface, i.e., *ErbB*, *RET*, and *PI3K-Akt* [4,31,38].

Trajectory	Gain	Perturbed Pathways	Loss	Perturbed Pathways
$C3\|P \rightarrow S3$	+2.83	Homologous recombination	−3.76	Hedgehog signaling pathway
	+2.41	Cell cycle	−3.45	Cytokine-cytokine receptor interaction
	+1.86	**ErbB signaling pathway**	−3.39	**PI3K-Akt signaling pathway**
	+1.10	cAMP signaling pathway	−3.15	TGF-beta signaling pathway
			−3.14	JAK-STAT signaling pathway
$S3 \rightarrow S1$	< 1.0	$\emptyset$	< 1.0	$\emptyset$
$S1 \rightarrow S2$	+1.36	cAMP signaling pathway	−1.28	JAK-STAT signaling pathway
	+1.18	**RET**	−1.22	Apoptosis
			−1.21	Cytokine-cytokine receptor interaction
			−1.12	Wnt signaling pathway
			−1.04	Notch signaling pathway
$S2 \rightarrow C1\|M$	+1.90	**RET**	−3.25	Wnt signaling pathway
	+1.59	PPAR signaling pathway	−3.11	JAK-STAT signaling pathway
			−2.77	Notch signaling pathway
			−2.48	Hedgehog signaling pathway
			−2.18	**PI3K-Akt signaling pathway**
$S2 \rightarrow C4\|M$	+4.48	Calcium signaling pathway	−3.06	p53 signaling pathway
	+4.17	cAMP signaling pathway	−2.74	Cell cycle
	+3.83	MAPK signaling pathway	−2.21	Homologous recombination
	+3.35	ECM-receptor interaction	−1.40	Apoptosis
	+3.20	Focal adhesion	−1.33	Cytokine-cytokine receptor interaction
$S1 \rightarrow C5$	+3.91	Cell cycle	−3.00	**RET**
	+3.17	p53 signaling pathway	−1.58	MAPK signaling pathway
	+2.85	Adherens junction	−1.41	cAMP signaling pathway
	+2.76	Cytokine-cytokine receptor interaction		
	+2.68	Wnt signaling pathway		
$S3 \rightarrow C2\|M$	+1.39	Homologous recombination	−3.65	TGF-beta signaling pathway
			−3.61	**PI3K-Akt signaling pathway**
			−3.34	ECM-receptor interaction
			−3.20	Focal adhesion
			−2.60	PPAR signaling pathway

Table A3: Perturbed pathways during the evolution of cell communities in primary and metastatic tumors (Fig. 3e Case 2). The top five perturbed pathways whose gain or loss greater than 1.0 along each edge of phylogeny are shown.

Trajectory	Gain	Perturbed Pathways	Loss	Perturbed Pathways
$C3\|P \rightarrow S1$	+2.83	Homologous recombination	−3.22	Hedgehog signaling pathway
	+2.47	Cell cycle	−3.10	TGF-beta signaling pathway
	+1.81	**ErbB signaling pathway**	−3.08	Cytokine-cytokine receptor interaction
	+1.02	cAMP signaling pathway	−2.93	**PI3K-Akt signaling pathway**
			−2.64	PPAR signaling pathway
$S1 \rightarrow S2$	+1.08	ECM-receptor interaction		
	+1.08	**ErbB signaling pathway**		
$S2 \rightarrow C4\|M$	+5.51	cAMP signaling pathway	−3.97	Cell cycle
	+5.12	Calcium signaling pathway	−3.83	p53 signaling pathway
	+4.45	MAPK signaling pathway	−3.20	Apoptosis
	+3.37	ECM-receptor interaction	−3.15	Cytokine-cytokine receptor interaction
	+3.08	**ErbB signaling pathway**	−3.00	Homologous recombination
$S2 \rightarrow C5$	+3.68	Cell cycle	−2.25	**RET**
	+3.18	p53 signaling pathway	−1.81	MAPK signaling pathway
	+2.50	Homologous recombination	−1.43	cAMP signaling pathway
	+2.16	Adherens junction	−1.24	Hedgehog signaling pathway
	+2.15	Cytokine-cytokine receptor interaction	−1.13	Calcium signaling pathway
$S1 \rightarrow C2\|M$	+1.39	Homologous recombination	−4.06	**PI3K-Akt signaling pathway**
			−3.70	TGF-beta signaling pathway
			−3.55	Focal adhesion
			−3.52	ECM-receptor interaction
			−2.87	Adherens junction

Table A4: Perturbed pathways during the evolution of cell communities in primary and metastatic tumors (Fig. 3e Case 3). The top five perturbed pathways whose gain or loss greater than 1.0 along each edge of phylogeny are shown.

Trajectory	Gain	Perturbed Pathways	Loss	Perturbed Pathways
$C3\|P \rightarrow S2$	+3.10	Cell cycle	−3.51	Hedgehog signaling pathway
	+3.10	**ErbB signaling pathway**	−2.41	Notch signaling pathway
	+2.93	Homologous recombination	−2.39	Cytokine-cytokine receptor interaction
	+1.70	cAMP signaling pathway	−2.34	JAK-STAT signaling pathway
	+1.66	HIF-1 signaling pathway	−2.07	Apoptosis
$S2 \rightarrow S1$	+1.62	cAMP signaling pathway	−2.02	Cytokine-cytokine receptor interaction
	+1.54	**RET**	−1.98	JAK-STAT signaling pathway
	+1.14	Calcium signaling pathway	−1.91	Apoptosis
			−1.75	Wnt signaling pathway
			−1.32	Cell cycle
$S1 \rightarrow C1\|M$	+1.85	**RET**	−3.52	Wnt signaling pathway
	+1.19	PPAR signaling pathway	−3.38	JAK-STAT signaling pathway
			−2.78	**PI3K-Akt signaling pathway**
			−2.76	Hedgehog signaling pathway
			−2.68	Notch signaling pathway
$S1 \rightarrow C4\|M$	+4.20	Calcium signaling pathway	−3.18	p53 signaling pathway
	+3.89	cAMP signaling pathway	−2.65	Cell cycle
	+3.40	MAPK signaling pathway	−1.99	Homologous recombination
	+2.76	Hedgehog signaling pathway	−1.64	Cytokine-cytokine receptor interaction
	+2.72	ECM-receptor interaction	−1.61	Apoptosis
$S2 \rightarrow C5$	+3.67	Cell cycle	−2.69	**RET**
	+2.76	Homologous recombination	−2.08	MAPK signaling pathway
	+2.56	p53 signaling pathway	−1.59	PPAR signaling pathway
	+1.85	mTOR signaling pathway	−1.43	cAMP signaling pathway
	+1.79	Adherens junction	−1.02	Hedgehog signaling pathway

Table A5: Perturbed pathways during the evolution of cell communities in primary and metastatic tumors (Fig. 3e Case 4). The top five perturbed pathways whose gain or loss greater than 1.0 along each edge of phylogeny are shown.

Trajectory	Gain	Perturbed Pathways	Loss	Perturbed Pathways
$C3\|P \rightarrow S1$	+2.38	Homologous recombination	−4.49	Cytokine-cytokine receptor interaction
	+1.56	**ErbB signaling pathway**	−4.23	**PI3K-Akt signaling pathway**
	+1.54	Cell cycle	−4.10	JAK-STAT signaling pathway
	+1.41	cAMP signaling pathway	−3.97	Hedgehog signaling pathway
			−3.74	Apoptosis
$S1 \rightarrow S2$	+1.89	cAMP signaling pathway	−1.66	Notch signaling pathway
	+1.69	**ErbB signaling pathway**	−1.27	JAK-STAT signaling pathway
	+1.47	HIF-1 signaling pathway	−1.14	Apoptosis
	+1.47	ECM-receptor interaction	−1.01	Cytokine-cytokine receptor interaction
	+1.43	Calcium signaling pathway		
$S2 \rightarrow C1\|M$	+1.43	PPAR signaling pathway	−2.53	Notch signaling pathway
	+1.19	**RET**	−2.44	Wnt signaling pathway
	+1.09	p53 signaling pathway	−2.35	Hedgehog signaling pathway
			−2.32	JAK-STAT signaling pathway
			−1.66	VEGF signaling pathway
$S2 \rightarrow C4\|M$	+4.40	Calcium signaling pathway	−2.37	p53 signaling pathway
	+3.91	cAMP signaling pathway	−1.93	Cell cycle
	+3.81	ECM-receptor interaction	−1.74	Homologous recombination
	+3.64	MAPK signaling pathway		
	+3.62	Focal adhesion		
$S1 \rightarrow C2\|M$	+1.84	Homologous recombination	−3.07	TGF-beta signaling pathway
	+1.39	Cell cycle	−2.77	**PI3K-Akt signaling pathway**
			−2.69	ECM-receptor interaction
			−2.59	Focal adhesion
			−2.58	PPAR signaling pathway