Supplementary Materials for "DINGO: Differential Network Analysis in Genomics"

In the following sections, we present detailed information about our method, DINGO, including additional simulation studies and results from the analysis of The Cancer Genome Atlas (TCGA) glioblastoma (GBM) data.

Section S1 DINGO model

As part of a detailed description of our DINGO model, we briefly review the covariance regression model by Hoff and Niu (2012) and generalize the model for the precision regression model, along with a simplified example for three genes.

Suppose that we observe gene expression profiles for p genes for a sample, denoted by a p-dimensional vector $\mathbf{y} \in \mathbb{R}^p$. For the sample, we also observe a covariate vector $\mathbf{x} \in \mathbb{R}^q$. In this paper, we consider a binary covariate, which represents the group of long-term survivors (LTSs) and the group of short-term survivors (STSs). The binary covariate is coded as $\mathbf{x} = (1, 1)^{\mathrm{T}}$ or $\mathbf{x} = (1, -1)^{\mathrm{T}}$ for the LTSs and STSs, respectively. Although we consider a binary covariate, the model can be easily generalized to incorporate multiple categories (such as multiple stages of disease and multiple subtypes) as well as continuous covariates (such as age and time). We now assume a general covariate vector $\mathbf{x} \in \mathbb{R}^q$.

Section S1.1 Covariance regression model

The covariance regression model (Hoff and Niu, 2012) is used to estimate a covariance function $\texttt{Cov}(\mathbf{y}|\mathbf{x})$ for covariates $\mathbf{x}$. The linear regression model expresses the conditional mean $\mathbb{E}(\mathbf{y}|\mathbf{x})$ as $\mathbf{Q}'\mathbf{x}$, where $\mathbf{Q}'$ is the $p \times q$ matrix of coefficients. This model restricts the p-dimensional vector of the conditional mean of $\mathbf{y}$ given $\mathbf{x}$ to a q-dimensional subspace of $\mathbb{R}^p$. Hoff and Niu (2012) suggested a covariance regression

model as a natural generalization of the mean regression to a model for covariance matrices

$$\texttt{Cov}(\mathbf{y}|\mathbf{x}) = \mathbf{\Psi}' + \mathbf{Q}'\mathbf{x}\mathbf{x}^{\mathrm{T}}\mathbf{Q}'^{\mathrm{T}},$$

where $\mathbf{\Psi}'$ is a $p \times p$ positive definite matrix and $\mathbf{Q}'$ is a $p \times q$ matrix. This covariance function is positive definite for all $\mathbf{x}$. They call $\mathbf{\Psi}'$ a *baseline* covariance matrix. In the covariance regression model, the covariance matrix of $\mathbf{y}$ is expressed as a baseline covariance matrix plus a rank-1, $p \times p$ positive definite matrix that depends on $\mathbf{x}$. This model describes marginal correlations among genes given covariates. In the DINGO model, we are interested in fitting partial correlations (correlations between any two genes when all the other genes remain constant) and visualizing a differential network that describes the differences in partial correlations between groups.

Section S1.2 Precision regression model

Our DINGO model is based on the mean function $\mathbb{E}(\mathbf{y}|\mathbf{x}) = \boldsymbol{\mathcal{G}}\mathbf{y} + \mathbf{Q}\mathbf{x}$, where $\boldsymbol{\mathcal{G}}$ is a $p \times p$ matrix of coefficients specifying relations among the p genes of $\mathbf{y}$, and $\mathbf{Q}$ is a $p \times q$ matrix of coefficients for $\mathbf{x}$. Therefore, we assume that each variable of $\mathbf{y}$ is affected by other $p-1$ genes of $\mathbf{y}$ as well as the explanatory variables in $\mathbf{x}$. We define the coefficient $\boldsymbol{\mathcal{G}}$ as a *global* component because it specifies relations among the p genes and is independent of $\mathbf{x}$. If $\mathbf{x}$ specifies groups of the GBM samples, LTSs and STSs, then $\boldsymbol{\mathcal{G}}$ reflects the global relations among the p genes of GBM patients regardless of the patients' survival times. We estimate the group-specific partial correlations in two steps. Step 1. Estimate the global relations. Step 2. Using the residual data after removing the effects of global relations, fit the covariance regression model (Hoff and Niu, 2012).

In Step 1 of our DINGO estimation, we consider the *global network model*,

$$\mathbf{y} = \boldsymbol{\mathcal{G}}\mathbf{y} + \boldsymbol{\epsilon} \tag{1}$$

where the elements of $\boldsymbol{\mathcal{G}} = (\mathcal{G}_{ab})_{p\times p}$ specify the global relations among the variables of $\mathbf{y}$, $\boldsymbol{\epsilon}$ is a $p \times 1$ vector following $N_p(\mathbf{0}, \boldsymbol{\mathcal{L}})$ where $\boldsymbol{\mathcal{L}}$ is the "local" Gaussian graphical model (GGM) whose elements specify relations among variables in $\mathbf{y}$ after removing the effects of the global relations. In Step 2 of the DINGO estimation, we connect the local GGM $\boldsymbol{\mathcal{L}}$ with the covariate $\mathbf{x}$ using the covariance regression model,

$$\begin{aligned} \boldsymbol{\mathcal{L}}(\mathbf{x})^{-1} &= \texttt{Cov}\,(\boldsymbol{\epsilon}|\mathbf{x}) \\ &= \mathbf{Q}\mathbf{x}\mathbf{x}^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}} + \boldsymbol{\Psi}, \end{aligned}$$

where $\mathbf{Q} = [Q_{ab}]_{p\times q}$ is a $p \times q$ matrix, which is the main construction of interest, and $\boldsymbol{\Psi} = \texttt{diag}(\psi_1, \ldots, \psi_p)$, with $\psi_1 > 0, \ldots, \psi_p > 0$, is a $p \times p$ diagonal matrix whose elements represent variances for pure noise in $\mathbf{y}$. From this model, we assume that after taking out the effects of relations among $\mathbf{y}$ and the covariates $\mathbf{x}$, the p variables in $\mathbf{y}$ are independent by restricting $\boldsymbol{\Psi}$ to be a diagonal matrix. By using the inverse formula in Miller (1981), the covariance regression model on $\boldsymbol{\mathcal{L}}(\mathbf{x})^{-1}$ can be expressed on the precision matrix function $\boldsymbol{\mathcal{L}}(\mathbf{x})$ in equation (1) in Section 2.1. This is the *precision regression model* as opposed to the covariance regression model. This model represents covariate-dependent Gaussian graphical models. We can calculate the group-specific GGMs. From equation (1), we can see that

$$\begin{aligned} \boldsymbol{\mathcal{L}} &= \texttt{Cov}(\boldsymbol{\epsilon})^{-1} \\ &= (\mathbf{I} - \boldsymbol{\mathcal{G}})^{-\mathrm{T}}\texttt{Cov}(\mathbf{y})^{-1}(\mathbf{I} - \boldsymbol{\mathcal{G}})^{-1} \\ &= (\mathbf{I} - \boldsymbol{\mathcal{G}})^{-\mathrm{T}}\boldsymbol{\mathcal{N}}(\mathbf{I} - \boldsymbol{\mathcal{G}})^{-1}. \end{aligned}$$

Thus, we can obtain the group-specific GGMs by the convolution

$$\begin{aligned} \boldsymbol{\mathcal{N}}(\mathbf{x}) &= \boldsymbol{\mathcal{G}} \oplus \boldsymbol{\mathcal{L}}(\mathbf{x}), \\ &= (\mathbf{I} - \boldsymbol{\mathcal{G}})^{\mathrm{T}}\boldsymbol{\mathcal{L}}(\mathbf{x})(\mathbf{I} - \boldsymbol{\mathcal{G}}), \end{aligned}$$

where $\boldsymbol{\mathcal{L}}(\mathbf{x})$ is taken from equation (1) in Section 2.1. Each element of $\boldsymbol{\mathcal{L}}(\mathbf{x})$ can

be calculated for the two-group setting. $\boldsymbol{\mathcal{L}}(\mathbf{x})$ for $\mathbf{x}_{(i)} = (1,1)^{\mathrm{T}}$ and $\mathbf{x}_{(i)} = (1,-1)^{\mathrm{T}}$ in the precision regression model determine dependencies in the LTSs and STSs, respectively. We denote the local group-specific component $\boldsymbol{\mathcal{L}}(\mathbf{x})$ as $\boldsymbol{\mathcal{L}}\left((1,1)^{\mathrm{T}}\right) = [\mathcal{L}_{ab}^{(1)}]_{p\times p}$ for the LTSs and $\boldsymbol{\mathcal{L}}\left((1,-1)^{\mathrm{T}}\right) = [\mathcal{L}_{ab}^{(2)}]_{p\times p}$ for the STSs. The diagonal and off-diagonal elements are

$$\mathcal{L}_{aa}^{(1)} = \frac{1}{\psi_a} - \frac{(Q_{a1}+Q_{a2})^2}{(1+\kappa^{(1)})\,\psi_a^2},\ \mathcal{L}_{ab}^{(1)} = -\frac{(Q_{a1}+Q_{a2})(Q_{b1}+Q_{b2})}{(1+\kappa^{(1)})\,\psi_a\psi_b},$$

where $\kappa^{(1)} = \sum_{k=1}^{p}(Q_{k1}+Q_{k2})^2/\psi_k$, and

$$\mathcal{L}_{aa}^{(2)} = \frac{1}{\psi_a} - \frac{(Q_{a1}-Q_{a2})^2}{(1+\kappa^{(2)})\,\psi_a^2},\ \mathcal{L}_{ab}^{(2)} = -\frac{(Q_{a1}-Q_{a2})(Q_{b1}-Q_{b2})}{(1+\kappa^{(2)})\,\psi_a\psi_b},$$

where $\kappa^{(2)} = \sum_{k=1}^{p}(Q_{k1}-Q_{k2})^2/\psi_k$. Additionally, we can obtain a *baseline* component denoted by $\boldsymbol{\mathcal{L}}\left((1,0)^{\mathrm{T}}\right) = [\mathcal{L}_{ab}^{(0)}]_{p\times p}$,

$$\mathcal{L}_{aa}^{(0)} = \frac{1}{\psi_a} - \frac{Q_{a1}^2}{(1+\kappa^{(0)})\,\psi_a^2},\ \mathcal{L}_{ab}^{(0)} = -\frac{Q_{a1}Q_{b1}}{(1+\kappa^{(0)})\,\psi_a\psi_b},$$

where $\kappa^{(0)} = \sum_{k=1}^{p} Q_{k1}^2/\psi_k$. For an edge between Y_a and Y_b, the local group-specific components that differ from the baseline component are determined by the coefficients Q_{a2} and Q_{b2}. If both $Q_{a2} = 0$ and $Q_{b2} = 0$, then the edge intensities for both local group-specific components are the same as those of the baseline component. All parameters $\mathbf{Q}$ and $\boldsymbol{\Psi}$ are estimated by adapting the expectation-maximization algorithm described in Hoff and Niu (2012) by setting $\boldsymbol{\Psi}$ as a diagonal matrix.

Section S1.3 Simplified example

In Section 2.2, we describe the detailed estimation of our DINGO model when we have two groups. We take a simple example where $p = 3$, $n = 4$ for two groups and

suppose we have the data as follows:

$$\begin{aligned}
\mathbf{y}_{(1)} &= \begin{pmatrix} y_{11} & y_{12} & y_{13} \end{pmatrix}^{\mathrm{T}} \in \mathbb{R}^3 \quad , \mathbf{x}_{(1)} = \begin{pmatrix} 1 & -1 \end{pmatrix}^{\mathrm{T}} \in \mathbb{R}^2 \\
\mathbf{y}_{(2)} &= \begin{pmatrix} y_{21} & y_{22} & y_{23} \end{pmatrix}^{\mathrm{T}} \in \mathbb{R}^3 \quad , \mathbf{x}_{(2)} = \begin{pmatrix} 1 & -1 \end{pmatrix}^{\mathrm{T}} \in \mathbb{R}^2 \\
\mathbf{y}_{(3)} &= \begin{pmatrix} y_{31} & y_{32} & y_{33} \end{pmatrix}^{\mathrm{T}} \in \mathbb{R}^3 \quad , \mathbf{x}_{(3)} = \begin{pmatrix} 1 & 1 \end{pmatrix}^{\mathrm{T}} \in \mathbb{R}^2 \\
\mathbf{y}_{(4)} &= \begin{pmatrix} y_{41} & y_{42} & y_{43} \end{pmatrix}^{\mathrm{T}} \in \mathbb{R}^3 \quad , \mathbf{x}_{(4)} = \begin{pmatrix} 1 & 1 \end{pmatrix}^{\mathrm{T}} \in \mathbb{R}^2.
\end{aligned}$$

Let $\mathbf{Y} = \begin{pmatrix} \mathbf{y}_{(1)} & \mathbf{y}_{(2)} & \mathbf{y}_{(3)} & \mathbf{y}_{(4)} \end{pmatrix}^{\mathrm{T}}$ and $\mathbf{X} = \begin{pmatrix} \mathbf{x}_{(1)} & \mathbf{x}_{(2)} & \mathbf{x}_{(3)} & \mathbf{x}_{(4)} \end{pmatrix}^{\mathrm{T}}$ be the data matrices for all four individuals. Suppose that we obtain the precision matrix, $\boldsymbol{\mathcal{N}}$ for the three variables from GLasso (Friedman et al., 2008) using $\mathbf{Y}$ as

$$\boldsymbol{\mathcal{N}} = \begin{pmatrix} 1 & 0.2 & 0 \\ 0.2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Then we have the global network model for individual i,

$$\begin{aligned}
\mathbf{y}_{(i)} &= \boldsymbol{\mathcal{G}}\mathbf{y}_{(i)} + \boldsymbol{\epsilon}_{(i)} \\
&= \begin{pmatrix} 0 & 0.2 & 0 \\ 0.2 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \mathbf{y}_{(i)} + \boldsymbol{\epsilon}_{(i)},
\end{aligned}$$

and this induces the equations,

$$\begin{aligned}
y_{i1} &= 0.2y_{i2} + \epsilon_{i1} \\
y_{i2} &= 0.2y_{i1} + \epsilon_{i2} \\
y_{i3} &= \epsilon_{i3}
\end{aligned}$$

for $i \in \{1, 2, 3, 4\}$. From the global network model for data matrix $\mathbf{Y}$, $\mathbf{Y} = \mathbf{Y}\boldsymbol{\mathcal{G}}^{\mathrm{T}} + \boldsymbol{\mathcal{E}}$, we can obtain the residual data matrix $\boldsymbol{\mathcal{E}} = \mathbf{Y}(\mathbf{I} - \boldsymbol{\mathcal{G}}^{\mathrm{T}}) = (\boldsymbol{\epsilon}_{(1)}, \ldots \boldsymbol{\epsilon}_{(n)})^{\mathrm{T}}$, with $\boldsymbol{\epsilon}_{(i)} \sim N_p(\mathbf{0}, \boldsymbol{\mathcal{L}})$ for all $i \in \{1, \ldots, n\}$.

In Step 2 of our DINGO method, we model the local GGM $\boldsymbol{\mathcal{L}}$ to be group-

specific by fitting the precision regression model. Let the parameters in the precision regression model in equation (1) of Section 2.1 be

$$\mathbf{\Psi} = \begin{pmatrix} \psi_1 & 0 & 0 \\ 0 & \psi_2 & 0 \\ 0 & 0 & \psi_3 \end{pmatrix}, \text{ and } \mathbf{Q} = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \\ Q_{31} & Q_{32} \end{pmatrix}.$$

Those matrices are obtained by using the EM algorithm described in Hoff and Niu (2012). Then we can obtain the local group-specific component $\boldsymbol{\mathcal{L}}(\mathbf{x})$ for the two the groups that have respective covariate vectors $\mathbf{x} = \begin{pmatrix} 1 & 1 \end{pmatrix}^{\mathrm{T}}$ and $\mathbf{x} = \begin{pmatrix} 1 & -1 \end{pmatrix}^{\mathrm{T}}$, by applying the matrices into equation (1) in Section 2.1. Or, each element of the local group-specific components are obtained by using equations listed in Section S1.2. By convolution in equation (2) in Section 2.1, we can build the group-specific GGMs. Then we make inference on the edge-wise differences of the sets of partial correlations obtained from the group-specific GGMs (described in Section 2.2 and Section S1.4).

Section S1.4 Bootstrap procedure to obtain differential scores

We build a differential network by thresholding the difference in the conditional dependencies between two groups for each pair of genes. The two group-specific GGMs are denoted by $\boldsymbol{\mathcal{N}}^{(1)}$ and $\boldsymbol{\mathcal{N}}^{(2)}$ for the LTSs and STSs. The corresponding sets of $p(p-1)/2$ partial correlations for the LTSs and STSs are denoted by $\{\hat{\rho}_{ab}^{(1)} : a, b \in V \text{ and } a < b\}$ and $\{\hat{\rho}_{ab}^{(2)} : a, b \in V \text{ and } a < b\}$, respectively. For an edge between a and b, we hypothesize that two conditional dependencies corresponding to a pair of genes a and b are the same: $H_0 : \rho_{ab}^{(1)} = \rho_{ab}^{(2)}$ vs. $H_A : \rho_{ab}^{(1)} \neq \rho_{ab}^{(2)}$. Given a conditional dependence estimate $\hat{\rho}$, the test statistic can be constructed by Fisher's Z transformation $\phi(\hat{\rho}) = 0.5 \log\{(1+\hat{\rho})/(1-\hat{\rho})\}$, and all the transformed conditional dependencies are denoted by $\{\phi_{ab}^{(1)} : a, b \in V \text{ and } a < b\}$ and $\{\phi_{ab}^{(2)} : a, b \in V \text{ and } a < b\}$, respectively. Specifically, one may reject the null hypothesis at level α if $|\phi_{ab}^{(1)} - \phi_{ab}^{(2)}| / \sqrt{1/(n_1 - p - 1) + 1/(n_2 - p - 1)} > \Phi^{-1}(1 - \alpha/2)$, where n_1 and n_2 are the sample sizes for the LTSs and STSs, and Φ is the cdf of $N(0, 1)$.

Although this statistic is asymptotic and follows a normal distribution under the null hypothesis, it is only valid when the sample sizes are large enough. Moreover, our group-specific GGMs are jointly estimated through the two components $\boldsymbol{\mathcal{G}}$ and $\boldsymbol{\mathcal{L}}(\mathbf{x})$ with parameters $\mathbf{Q}$ and $\boldsymbol{\Psi}$ from all $n_1 + n_2 = n$ samples. The two conditional dependencies are correlated and the asymptotic variance of the difference is not known. The bootstrap estimate of the standard error for the difference $\hat{\phi}_{ab}^{(1)} - \hat{\phi}_{ab}^{(2)}$ is calculated as follows:

1. Obtain $\hat{\boldsymbol{\mathcal{E}}} = \mathbf{Y}(\mathbf{I} - \hat{\boldsymbol{\mathcal{G}}}^{\mathrm{T}})$, where $\hat{\boldsymbol{\mathcal{G}}}$ is the estimate of $\boldsymbol{\mathcal{G}}$ from Step 1 of DINGO.
2. Draw the b^{th} bootstrap resample with replacement to obtain $\hat{\boldsymbol{\mathcal{E}}}^{*b}$ and $\mathbf{X}^{*b}$.
3. Using $\hat{\boldsymbol{\mathcal{E}}}^{*b}$ and $\mathbf{X}^{*b}$, fit the precision regression model in equation (1) in Section 2.1 and calculate Fisher's Z transformed conditional dependence differences $d_{ij}^b = \hat{\phi}_{ij}^{(1)b} - \hat{\phi}_{ij}^{(2)b}$ for all $i < j$ and $i, j \in V$.

Iterate step 2 through step 3 above B times, and calculate the bootstrap standard error estimates for all edges $i < j$ as $s_{ij}^B = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}\left(d_{ij}^b - \bar{d}_{ij}\right)^2}$, where $\bar{d}_{ij} = \sum_{b=1}^{B} d_{ij}^b/B$. From this bootstrap procedure, we construct a *differential score* as

$$\textit{Differential Score:} \quad \delta_{ij}^{(12)} = \frac{\hat{\phi}_{ij}^{(1)} - \hat{\phi}_{ij}^{(2)}}{s_{ij}^B}.$$

Section S2 Application

Section S2.1 Long-term survivors and short-term survivors from TCGA glioblastoma data

Among 233 patients (TCGA GBM data), 70 were censored; the quantiles of the survival times (in days) were 5 (0%),153.50 (25%), 219.22 (33%), 341.10 (45%), 383.00 (50%), 407.20 (55%) 481.76 (66%), 541.50 (75%) and 3880 (100%). Due to censoring, the survival time T is not always observable: instead, for patient i, we

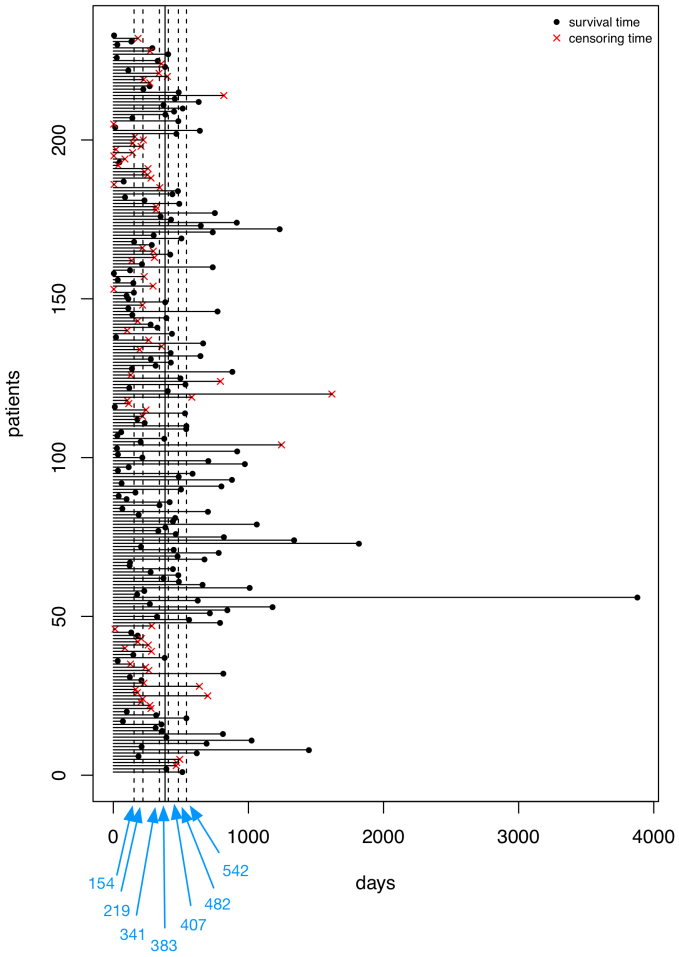

Figure S1: Survival and censoring time distribution. The vertical lines, from left to right, are 25% (dashed), 33% (dashed), 45% (dashed), 50% (solid), 55% (dashed), 66% (dashed) and 75% (dashed).

observed

$$z_i = \min(t_i, c_i) \text{ and } \delta_i = I(t_i \leq c_i), \tag{2}$$

where c_i is the censoring time for an individual i. For the cutoffs k_1 and k_2, we define $x_i = 1$ if $z_i > k_2$, $x_i = -1$ if $z_i < k_1$ and $\delta_i = 1$. The cutoffs k_1 and k_2 are used to discretize the total dataset into distinct extreme survival groups: LTSs versus STSs.

Table S1: The numbers of patients (%) in the survival groups according to the cutoffs k_1 and k_2 set by the quantiles of the survival times.

k_1	k_2	LTSs (%)	STSs (%)	Missed (%)
383.00 (50 %)	383.00 (50%)	92 (39.48)	81 (34.76)	60 (25.75)
341.10 (45%)	407.20 (55%)	83 (35.62)	73 (31.33)	77 (33.05)
219.22 (33%)	481.76 (66%)	64 (27.47)	54 (23.18)	115 (49.36)
153.50 (25%)	541.50 (75%)	48 (20.60)	41 (17.60)	144 (61.80)

The middle group ($k_1 \leq x_i \leq k_2$) and the patients with censoring time less than k_1 were not used in this analysis.

Figure S1 displays the survival/censoring time for all 233 patients. Vertical lines depict the quantiles of the observed survival times. For several choices of the cutoffs k_1 and k_2, Table S1 shows the number of patients (%) classified into each group. We chose $k_1 = 341.10$ and $k_2 = 407.20$, which correspond to the $45th$ and $55th$ percentiles, and result in 83 patients being classified as LTSs and 73 classified as STSs, with 77 patients being excluded from the analysis.

Section S2.2 Analysis of RTK/PI3K, p53, Rb signaling pathways

For the 83 LTSs and 73 STSs, we analyzed mRNA, DNA copy number, methylation and microRNA data corresponding to genes involved in the GBM pathways `http://cbio.mskcc.org/cancergenomics/gbm/pathways`. Those genes are listed as follows: AKT1, AKT2, AKT3, PIK3CA, PIK3CB, PIK3CD, PIK3R1, PIK3R2, FOXO1, FOXO3, FOXO4, PIK3CG, PDPK1, IRS1, SRC, GAB1, PTEN, IGF1R, PDGFRA, PDGFRB, EGFR, ERBB2, ERBB3, FGFR1, FGFR2, MET, NRAS, HRAS, KRAS, ARAF, BRAF, RAF1, GRB2, NF1, CBL, SPRY2, CDKN2A, CDKN2C, CDKN2B, CCND1, CCND2, CDK4, CDK6, RB1, MDM2, MDM4, TP53, PIK3C2B, PIK3C2G.

Data processing

We downloaded data from TCGA website for analysis, which included mRNA, DNA copy number, methylation and microRNA data generated from the Affymetrix HT Human Genome U133 Array Plate Set, Agilent Human Genome CGH Microarray 244A, Illumina Infinium Human DNA Methylation 27, and the Agilent 8×15K Human miRNA-specific microarray. For DNA copy number and methylation data, we took the first principal components for several sites that correspond to a gene in the GBM pathways. The proportion of variances explained by the first principal

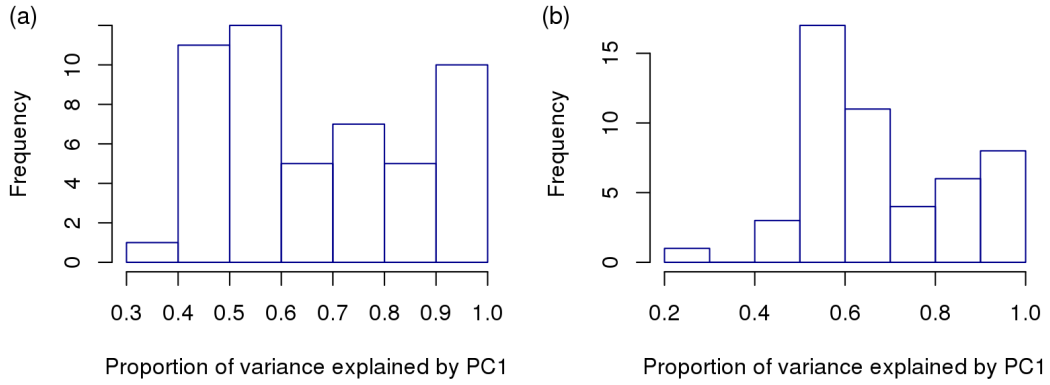

Figure S2: Proportion of variances explained by the first principal component. (a) Copy number data (b) Methylation data.

components are displayed in Figure S2. For DNA copy number data, the quantiles corresponding to probabilities 0, 0.25, 0.5, 0.75 and 1 of the numbers of sites corresponding to genes were 1, 4, 7, 12 and 43, and the quantiles of the proportions of variances explained by the first principal components were 0.35, 0.51, 0.63, 0.85 and 1. For methylation data, the quantiles corresponding to probabilities 0, 0.25, 0.5, 0.75 and 1 of the numbers of sites corresponding to genes were 1, 2, 2, 2 and 21 and the quantiles of the proportions of variances explained by the first principal components were 0.29, 0.55, 0.64, 0.82 and 1. For microRNA data, we took only human microRNAs. The number of vertices (p) consisted of 49 genes for mRNA, 51 for copy number, 50 for methylation, and 470 for microRNA.

Section S2.3 Comprehensive analysis of pathways in KEGG, BIORCARTA and REACTOME

We undertook a more comprehensive analysis of multiple pathways curated from existing databases (such as KEGG, BIOCARTA and REACTOME). This provides a more comprehensive landscape of GBM progression. We analyzed 24 pathways from KEGG and BIOCARTA, and 70 pathways from REACTOME, which we selected on the basis of the literature Hanahan and Weinberg (2011), McLendon et al. (2008), Parsons et al. (2008) and Verhaak et al. (2010). The LTSs and STSs are defined in Section S2.1 (83 LTSs and 73 STSs).

Using mRNA expression data, we performed DINGO for the genes in each pathway. From this analysis, we can investigate which pathways are differentially expressed between LTSs and STSs. For our DINGO method, we set the cutoff for the absolute values of the differential scores as 2: edges with greater than 2 in their absolute values of the differential scores are determined to be edges in the differential networks. Pathways that possess more differential edges are considered to better differentiate between the two groups, LTSs and STSs.

Table S2, Table S3 and Table S4 display the names of pathways, numbers of genes per pathway, the numbers (proportions) of the differential edges from DINGO, and lists (degree) of hub genes. The hub genes are defined by vertices that have degrees greater than 5% of the total number of differential edges. By applying our DINGO model to genes in the pathways from KEGG and BIOCARTA, we found the p53 pathway to best differentiate between the groups. The "pathways in cancer" was ranked in the top 2, and pathways related to glioblastoma and apoptosis had more than 4% differential edges. From REACTOME, the GRB2 and NOTCH2 pathways were the top 2 ranked pathways with more than 7% differential edges. We also found FGFR-related pathways to be top-ranked pathways with more than 5% differential

edges. For all three databases, KEGG, BIOCARTA and REACTOME, the apoptosis pathways were among those that best differentiated between LTSs and STSs.

Biological interpretation

The p53 gene has an important role in promoting apoptosis in response to several oncogenes such as the c-Myc oncogene and CASP9 gene (the top hub in the pathway with 19 differential edges), which is an essential downstream component of p53 in Myc-induced apoptosis (Soengas et al., 1999). Moreover, p53 mutations are one of the genetic events that molecularly differentiate clinical subtypes of GBM (Ruano et al., 2009). CASP9 with APAF1 enhances p53-mediated apoptotic cell death in glioblastoma (Shinoura et al., 2002). The pathways in cancer that ranked among the top 2 from KEGG (`http://www.genome.jp/kegg-bin/show_pathway?hsa05200`) are integrated pathways for 14 cancers, including glioblastoma. The PIK3R1 genes (with degree 106) in the cancer pathways was reported as a potential therapeutic target in GBM (Weber et al., 2011). We found the integrin signaling pathways to be among the top-ranked pathways from REACTOME. From GRB2 to the MAPK signaling pathway was the top-ranked pathway, and from P130CAS to the MAPK signaling pathway was the 12th ranked pathway among 70 pathways from REACTOME (Table S3). The integrin signaling pathway, which explains how MAPK signaling is activated by GRB2 and P130CAS, is displayed in Box 1 in (Guo and Giancotti, 2004). Integrins are expressed in glioblastoma cells, have a possible role in invasion, and are potential treatment targets for glioblastoma (Desgrosellier and Cheresh, 2010; Bello et al., 2001). It was also reported that MAPK signaling contributes to the development of malignant glioblastoma (Nakada et al., 2011; Sheng et al., 2010). Because FGFR inhibition can reduce proliferation and induce cell death in tumor models, FGFRs are considered to be attractive targets for therapeutic intervention in cancer (Dienstmann et al., 2014). We found FGFR pathways to rank as numbers 3, 4, 5 and 7, with more than 5% differential edges (Table S3). The administration of an FGFR prolongs the survival of mice that harbor FGFR3-

TACC3-initiated glioblastomas (Dienstmann et al., 2014; Singh et al., 2012). It is reported that increased NOTCH2 signaling (rank 2 from REACTOME in Table S3) contributes to increased tumor growth in GBM (Fan et al., 2010; Tchorz et al., 2012).

Effects of c-Myc gene

For assessing the direct effects of the c-Myc gene, this comprehensive analysis of pathways included the MYC gene. The differential networks for neighbors of the MYC gene are shown in Figure S3.

Table S2: Differential networks for genes in pathways from KEGG and BIOCARTA

	Pathway	p[a]	DINGO[b]	Hub genes (degree)[c]
1	P53 PATHWAY	64	105 (0.05)	BAI1 (10), BCL2 (14), CASP9 (19), CCNB2 (6), CCND1 (7), CCND2 (9), CCND3 (7), CD82 (12), CDKN1A (7), CHEK1 (10), E2F1 (11), FAS (6), IGF1 (14), PCNA (11), SERPINE1 (18)
2	PATHWAYS IN CANCER	301	2110 (0.05)	CEBPA (117), PIK3R1 (106)
3	ERK5 PATHWAY	17	6 (0.04)	CREB1 (5), MAPK1 (1), MAPK3 (1), MAPK7 (1), MEF2A (1), MEF2C (2), SHC1 (1)
4	IGF1MTOR PATHWAY	19	7 (0.04)	AKT1 (2), EIF2S1 (5), EIF2S3 (1), PDPK1 (1), PPP2CA (2), RPS6 (2), RPS6KB1 (1)
5	MAPK PATHWAY	254	1310 (0.04)	CACNA2D2 (66), CEBPA (97), FGFR1 (74), PPP3CA (92), TP53 (67)
6	APOPTOSIS	87	147 (0.04)	BAD (9), BCL2 (8), BID (9), BIRC3 (12), CASP7 (9), CCR5 (8), CD247 (15), CD4 (13), CSF2RB (31), DFFA (33), IL1A (11), IL1B (15), PPP3CA (8), PRKAR1A (11)
7	P53HYPOXIA PATHWAY	22	9 (0.04)	ABCB1 (2), ATM (3), BAX (2), CSNK1A1 (3), HIC1 (4), IGFBP3 (2), TAF1 (2)
8	TGF BETA PATHWAY	83	132 (0.04)	BMP2 (11), BMP4 (14), BMP5 (9), BMPR1A (9), ID1 (10), MAP3K7 (25), MAPK1 (20), RBX1 (12), RHOA (9), ROCK2 (18), SMAD2 (8), SMAD4 (8), SMAD9 (11), TGFB1 (12), THBS3 (9), ZFYVE16 (14)
9	VEGF PATHWAY	85	133 (0.04)	ARNT (16), FLT4 (18), HSPB1 (7), MAP2K2 (11), MAPK11 (7), MAPKAPK2 (14), NRAS (8), PIK3CA (10), PIK3CB (9), PIK3R3 (9), PLA2G3 (15), PLA2G4B (26), PLA2G5 (11), PPP3CB (8), SHC2 (9), SPHK2 (14), VEGFA (10)
10	EGF PATHWAY	29	15 (0.04)	CSNK2A1 (9), EGF (3), EGFR (1), FOS (1), JAK1 (2), MAP2K1 (1), MAPK3 (1), PIK3CA (2), PLCG1 (2), PRKCA (1), RASA1 (3), STAT3 (3), STAT5B (1)
11	GLIOMA	59	61 (0.04)	CALM1 (4), CDK4 (7), CDKN2A (15), HRAS (6), IGF1 (4), NRAS (4), PIK3CG (11), PIK3R1 (6), PLCG2 (12), SOS2 (15), TP53 (4)
12	PTEN PATHWAY	16	4 (0.03)	FOXO3 (1), ITGB1 (1), MAPK1 (1), PDPK1 (3), PIK3CA (2)
13	P38MAPK PATHWAY	39	24 (0.03)	DAXX (8), MAP3K5 (2), MAP3K7 (4), MAPKAPK2 (4), MEF2A (2), RAPGEF2 (3), RIPK1 (9), SHC1 (3), STAT1 (2), TGFB1 (2), TRADD (2)
14	NOTCH PATHWAY	36	20 (0.03)	APH1A (4), DLL3 (5), DVL1 (2), DVL3 (2), HDAC1 (2), JAG1 (2), LFNG (3), NCOR2 (2), NOTCH1 (4), NOTCH3 (5), NUMB (2), PSENEN (4)
15	ERBB PATHWAY	80	100 (0.03)	CAMK2A (18), CAMK2G (6), CBLC (8), EIF4EBP1 (6), MAPK10 (12), NCK1 (10), NCK2 (18), PAK1 (7), PAK6 (18), PAK7 (18)
16	IL1R PATHWAY	31	14 (0.03)	IL1A (3), IRAK1 (1), JUN (2), MAP3K1 (1), MAP3K7 (1), MAPK14 (1), MYD88 (2), NFKB1 (2), RELA (7), TGFB3 (1), TOLLIP (6), TRAF6 (1)
17	MTOR PATHWAY	56	45 (0.03)	BRAF (4), EIF4A2 (3), EIF4E2 (12), EIF4EBP1 (8), IGF1 (6), MAPK1 (3), PIK3CA (3), PPP2CA (4), RPS6 (9), RPS6KA1 (5), RPS6KA3 (7), RPS6KB2 (4), VEGFB (7)
18	CDK5 PATHWAY	10	1 (0.02)	HRAS (1), MAPK3 (1)
19	RAS PATHWAY	23	5 (0.02)	CDC42 (4), PIK3CA (2), PIK3R1 (1), RAF1 (1), RALGDS (1), RELA (1)
20	EGFR SMRTE PATHWAY	11	1 (0.02)	MAP3K1 (1), ZBTB16 (1)
21	ERK PATHWAY	26	5 (0.02)	ELK1 (1), MKNK2 (1), NGFR (2), PPP2CA (1), PTPRR (3), STAT3 (2)
22	BARR MAPK PATHWAY	12	1 (0.02)	MAPK1 (1), PLCB1 (1)
23	RB PATHWAY	12	1 (0.02)	CHEK1 (1), MAPK14 (1)
24	MET PATHWAY	34	6 (0.01)	CRKL (1), GAB1 (1), GRB2 (5), HRAS (1), ITGB1 (1), MAP2K2 (1), PAK1 (1), PIK3R1 (1)

[a] Number of genes in the pathway (p).

[b] Number (proportion) of differential edges from DINGO. The proportions are divided by the number of all possible pairs of vertices ($p(p-1)/2$).

[c] Hub genes are defined by vertices that have degrees greater than 5% of the total number of differential edges.

Table S3: Differential networks for genes in pathways from REACTOME

	Pathway	p[a]	DINGO[b]	Hub genes (degree)[c]
1	GRB2 SOS PROVIDES LINKAGE TO MAPK SIGNALING FOR INTERGRINS	13	6 (0.08)	APBB1IP (2), PTK2 (2), RAP1A (3), SRC (3), TLN1 (2)
2	SIGNALING BY NOTCH2	10	3 (0.07)	ADAM10 (2), APH1A (2), NCSTN (1), PSEN1 (1)
3	SIGNALING BY FGFR3 MUTANTS	10	3 (0.07)	FGF1 (1), FGF18 (1), FGF2 (2), FGF5 (2)
4	FGFR4 LIGAND BINDING AND ACTIVATION	10	3 (0.07)	FGF1 (2), FGF2 (2), FGF20 (1), FGFR4 (1)
5	FGFR LIGAND BINDING AND ACTIVATION	18	10 (0.07)	FGF1 (1), FGF22 (1), FGF23 (1), FGF3 (9), FGF6 (2), FGF8 (1), FGF9 (2), FGFR1 (1), FGFR3 (1), FGFR4 (1)
6	TGF BETA RECEPTOR SIGNALING IN EMT EPITHELIAL TO MESENCHYMAL TRANSITION	12	4 (0.06)	ARHGEF18 (1), PARD3 (1), RHOA (1), RPS27A (4), UBA52 (1)
7	FGFR2C LIGAND BINDING AND ACTIVATION	11	3 (0.05)	FGF17 (1), FGF2 (3), FGF8 (1), FGFR2 (1)
8	APOPTOSIS	132	448 (0.05)	BAD (36), BID (32), CASP9 (32), CYCS (28), DAPK2 (34), E2F1 (33), HIST1H1B (26), MAPK8 (34), PKP1 (49), PRKCD (30), PSMA7 (31), PSMC1 (31)
9	PROLONGED ERK ACTIVATION EVENTS	17	7 (0.05)	BRAF (1), HRAS (1), KIDINS220 (7), MAPK1 (1), MAPK3 (1), NTRK1 (1), RAF1 (1), RAP1A (1)
10	EXTRINSIC PATHWAY FOR APOPTOSIS	13	4 (0.05)	CASP10 (1), CFLAR (1), FAS (1), RIPK1 (3), TNFSF10 (2)
11	SHC1 EVENTS IN EGFR SIGNALING	13	4 (0.05)	EGFR (1), GRB2 (1), MAP2K2 (1), MAPK1 (1), RAF1 (3), YWHAB (1)
12	P130CAS LINKAGE TO MAPK SIGNALING FOR INTEGRINS	13	4 (0.05)	APBB1IP (1), FGA (1), PTK2 (2), RAP1A (2), SRC (2)
13	SIGNALING BY CONSTITUTIVELY ACTIVE EGFR	16	6 (0.05)	CBL (2), CDC37 (2), GAB1 (4), HRAS (1), PIK3CA (2), UBA52 (1)
14	PI3K AKT ACTIVATION	29	20 (0.05)	AKT1 (6), BAD (3), CASP9 (2), FOXO1 (5), IRS2 (5), NR4A1 (3), NTRK1 (2), PIK3CA (2), PIK3CB (3), PIK3R1 (4), TSC2 (4)
15	RAS ACTIVATION UOPN CA2 INFUX THROUGH NMDA RECEPTOR	15	5 (0.05)	CALM1 (2), CAMK2A (3), CAMK2B (3), GRIN1 (2)
16	SIGNALING BY EGFR IN CANCER	91	193 (0.05)	ADCY1 (14), ADCY7 (10), ADCY8 (31), ADCY9 (11), ADRBK1 (21), AKT1 (10), CDC37 (11), CDC42 (22), FOXO3 (14), ITPR2 (21), MAPK1 (11), MAPK3 (15), PDPK1 (10), PRKACG (10), PRKAR1B (19), PRKCA (17), SPRY1 (12), YWHAB (12)
17	DOWNREGULATION OF TGF BETA RECEPTOR SIGNALING	19	8 (0.05)	PPP1CA (1), PPP1CB (2), RPS27A (2), SMAD3 (2), SMAD7 (1), SMURF1 (1), SMURF2 (1), ZFYVE9 (6)
18	SIGNALING BY NOTCH4	10	2 (0.04)	ADAM10 (1), APH1A (2), PSEN1 (1)
19	TRAF6 MEDIATED INDUCTION OF NFKB AND MAP KINASES UPON TLR7 8 OR 9 ACTIVATION	65	92 (0.04)	ATF2 (17), CREB1 (15), DUSP6 (10), HMGB1 (12), MAP2K2 (15), MAP2K3 (5), MAP2K6 (5), MAPK8 (9), PPP2CA (5), PPP2R1A (16), PPP2R1B (5), RPS6KA1 (8), TLR7 (9), TRAF6 (5)
20	SIGNALING BY ERBB2	84	153 (0.04)	ADCY2 (17), ADRBK1 (17), CASP9 (11), CDC37 (26), CREB1 (8), EREG (26), FOXO3 (18), FOXO4 (11), FYN (9), MAP2K1 (9), MAPK1 (12), PLCG1 (20), PRKAR1B (18), YES1 (10)
21	NOTCH1 INTRACELLULAR DOMAIN REGULATES TRANSCRIPTION	33	23 (0.04)	CREBBP (20), HDAC4 (2), NCOR2 (2), TBL1XR1 (2), TLE1 (2), TLE3 (2)
22	SIGNALING BY TGF BETA RECEPTOR COMPLEX	54	62 (0.04)	CCNT1 (14), CCNT2 (11), FKBP1A (5), MEN1 (5), MTMR4 (4), PPP1CB (6), RHOA (5), SMURF2 (8), TGFBR2 (4), UBE2D3 (16), USP9X (5), XPO1 (11)
23	GASTRIN CREB SIGNALLING PATHWAY VIA PKC AND MAPK	154	509 (0.04)	EDN1 (47), EDN2 (30), EDN3 (32), GNA15 (46), GNB5 (34), GNG4 (40), HBEGF (27), HRH1 (33), NPFFR1 (48), RGS19 (33)
24	IL1 SIGNALING	32	21 (0.04)	CHUK (6), IKBKB (3), IL1A (8), IL1B (3), IL1RN (3), MAP2K4 (4), MAP3K7 (3), NOD1 (2), PELI1 (2), TRAF6 (2)
25	PI3K EVENTS IN ERBB4 SIGNALING	29	17 (0.04)	AKT2 (6), CDKN1A (2), CREB1 (3), ERBB4 (1), FOXO3 (2), PDPK1 (4), PIK3CA (3), PIK3R1 (1), PTEN (5), RPS6KB2 (1), TRIB3 (4), TSC2 (2)
26	NFKB AND MAP KINASES ACTIVATION MEDIATED BY TLR4 SIGNALING REPERTOIRE	62	79 (0.04)	APP (11), ATF1 (4), CREB1 (4), DUSP4 (15), DUSP6 (18), IRAK1 (5), LY96 (6), MAP2K2 (10), MAPKAPK3 (8), MEF2C (5), NFKBIB (11), PPP2R1A (5), RPS6KA1 (6), TLR3 (16)
27	SIGNALING BY FGFR	92	173 (0.04)	CALM2 (25), CHUK (33), FGF23 (10), FGF7 (12), FOXO3 (9), ITPR2 (18), PDE1B (15), PIK3R1 (11), PPP2CB (18), PRKACA (11), PRKAR1B (25), SRC (13), TSC2 (14)
28	GRB2 EVENTS IN ERBB2 SIGNALING	19	7 (0.04)	EGF (1), HRAS (7), MAPK1 (1), MAPK3 (1), NRAS (1), NRG1 (1), NRG2 (1), RAF1 (1)
29	PI3K EVENTS IN ERBB2 SIGNALING	35	24 (0.04)	AKT1 (3), CHUK (2), CREB1 (2), ERBB4 (8), GRB2 (2), HBEGF (2), PIK3R1 (11), PTEN (2), RPS6KB2 (5), TSC2 (3)
30	PI3K CASCADE	55	59 (0.04)	EIF4EBP1 (4), FGF18 (10), FGF2 (21), FGF20 (3), FGF5 (15), INSR (5), PIK3R2 (3), PPM1A (4), PRKAA1 (4), PRKAG2 (6), STK11 (6), TSC2 (6)
31	SIGNALING BY NOTCH	77	116 (0.04)	E2F3 (11), EIF2C1 (6), EIF2C3 (14), EIF2C4 (7), FBXW7 (8), LFNG (13), PSEN1 (16), RPS27A (7), ST3GAL6 (6), TBL1X (18), TBL1XR1 (18), TLE1 (16), TP53 (6), UBA52 (15)
32	MAP KINASE ACTIVATION IN TLR CASCADE	43	35 (0.04)	ATF1 (4), ATF2 (2), DUSP3 (4), DUSP4 (9), DUSP6 (9), MAP2K1 (4), MAP2K2 (6), MAPK10 (3), MEF2A (2), MEF2C (4), PPP2CB (3), PPP2R1A (12), RIPK2 (3)
33	PRE NOTCH EXPRESSION AND PROCESSING	31	18 (0.04)	ATP2A2 (1), B4GALT1 (2), CCND1 (2), CREBBP (1), E2F1 (1), E2F3 (7), EP300 (2), RAB6A (8), RBPJ (3), SNW1 (3), TFDP1 (1), TMED2 (3), TNRC6B (2)
34	DOWNSTREAM SIGNALING OF ACTIVATED FGFR	82	127 (0.04)	ADCY9 (10), FGF18 (9), FGF3 (10), FGF5 (7), FGF9 (27), GSK3A (8), ITPR2 (7), MAP2K2 (10), MAPK3 (9), PRKACA (17), PRKAR1B (24), PTEN (7), TSC2 (17)
35	PKB MEDIATED EVENTS	21	8 (0.04)	CAB39 (1), EIF4B (1), PDE3B (1), PPM1A (1), PRKAA1 (1), PRKAB2 (1), PRKAG2 (1), RPS6KB1 (8), TSC2 (1)

[a] Number of genes in the pathway (p).

[b] Number (proportion) of differential edges from DINGO. The proportions are divided by the number of all possible pairs of vertices ($p(p-1)/2$).

[c] Hub genes are defined by vertices that have degrees greater than 5% of the total number of differential edges.

Table S4: Differential networks for genes in pathways from REACTOME

	Pathway	p[a]	DINGO[b]	Hub genes (degree)[c]
36	MAPK TARGETS NUCLEAR EVENTS MEDIATED BY MAP KINASES	29	15 (0.04)	ATF1 (2), ATF2 (1), DUSP3 (2), DUSP4 (7), DUSP6 (8), FOS (2), MAPK1 (1), MEF2A (2), MEF2C (2), PPP2CB (2), RPS6KA5 (1)
37	ACTIVATED NOTCH1 TRANSMITS SIGNAL TO THE NUCLEUS	20	7 (0.04)	APH1A (2), DTX4 (2), NCSTN (1), NUMB (2), PSEN1 (3), RPS27A (1), UBA52 (3)
38	CDK MEDIATED PHOSPHORYLATION AND REMOVAL OF CDC6	45	36 (0.04)	CDC6 (7), CDK2 (11), PSMA3 (10), PSMA4 (2), PSMA6 (3), PSMB8 (7), PSMB9 (6), PSMC2 (2), PSMC6 (3), PSMD14 (3), PSMD8 (2), PSME1 (3), PSME2 (6), RPS27A (2)
39	SIGNALING BY ERBB4	76	102 (0.04)	BTRC (7), CREB1 (9), CSN2 (8), EREG (7), GHR (10), KRAS (10), MAPK1 (7), MDM2 (9), PIK3CA (7), PTEN (9), RAF1 (12), RBX1 (7), STAT5B (15), TSC2 (9), UBA52 (14), YAP1 (7)
40	ROLE OF DCC IN REGULATING APOPTOSIS	8	1 (0.04)	APPL1 (1), MAGED1 (1)
41	REGULATION OF APOPTOSIS	53	49 (0.04)	PSMA1 (20), PSMA3 (4), PSMB7 (5), PSMB8 (14), PSMB9 (13), PSMC5 (3), PSMD4 (5), PSMD6 (3), PSMD8 (3)
42	REGULATION OF HYPOXIA INDUCIBLE FACTOR HIF BY OXYGEN	22	8 (0.03)	CA9 (1), CREBBP (1), EGLN1 (1), EPO (1), TCEB2 (1), UBA52 (1), UBE2D2 (1), UBE2D3 (3), VEGFA (6)
43	EGFR DOWNREGULATION	22	8 (0.03)	AP2A2 (1), AP2S1 (1), CLTC (5), EGF (1), EPS15 (4), EPS15L1 (1), HGS (2), STAM (1)
44	SIGNALING BY FGFR IN DISEASE	102	178 (0.03)	CALM2 (27), CHUK (33), FGF7 (11), FGFR1OP (11), ITPR2 (29), PPP2CB (10), PRKAR1B (28), PRKCE (12), RPS27A (13), SRC (11), UBA52 (10)
45	P53 DEPENDENT G1 DNA DAMAGE RESPONSE	51	43 (0.03)	PSMA1 (5), PSMA3 (3), PSMB7 (3), PSMB8 (19), PSMB9 (20), PSMD4 (3)
46	ACTIVATED POINT MUTANTS OF FGFR2	14	3 (0.03)	FGF1 (2), FGF18 (1), FGF2 (1), FGF5 (2)
47	SIGNALING BY NOTCH1	51	42 (0.03)	ARRB2 (4), CCNC (8), DLK1 (6), HDAC2 (4), HDAC3 (4), HEY2 (7), NCOR1 (10), PSEN1 (8), SNW1 (3), TLE1 (7), TLE2 (4), TLE3 (5)
48	P53 INDEPENDENT G1 S DNA DAMAGE CHECKPOINT	47	35 (0.03)	CDC25A (2), PSMA1 (10), PSMA3 (5), PSMB2 (3), PSMB8 (4), PSMB9 (5), PSMC2 (3), PSMC3 (3), PSMC6 (3), PSMD1 (3), PSMD13 (3), PSMD14 (7), PSMD7 (3), PSMD8 (12)
49	ERK MAPK TARGETS	20	6 (0.03)	DUSP3 (1), DUSP6 (6), MAPK14 (1), MEF2A (1), MEF2C (1), RPS6KA1 (1), RPS6KA5 (1)
50	FGFR1 LIGAND BINDING AND ACTIVATION	12	2 (0.03)	FGF17 (1), FGF2 (2), FGF8 (1)
51	SHC1 EVENTS IN ERBB4 SIGNALING	17	4 (0.03)	HRAS (4), MAPK3 (1), NRG1 (1), NRG2 (1), RAF1 (1)
52	CD28 DEPENDENT PI3K AKT SIGNALING	17	4 (0.03)	CD28 (1), CD86 (4), LCK (1), PIK3CA (1), PIK3R1 (1)
53	G BETA GAMMA SIGNALLING THROUGH PI3KGAMMA	19	5 (0.03)	AKT2 (1), GNB1 (1), GNB3 (2), GNG13 (1), GNG7 (3), PIK3CG (1), PIK3R5 (1)
54	P38MAPK EVENTS	13	2 (0.03)	KRAS (1), RALB (2), RALGDS (1)
55	ACTIVATED TAK1 MEDIATES P38 MAPK ACTIVATION	13	2 (0.03)	MAPK11 (1), NOD2 (1), TRAF6 (2)
56	SIGNALING BY FGFR MUTANTS	36	16 (0.03)	BCR (1), CPSF6 (1), FGF1 (1), FGF18 (2), FGF2 (2), FGF20 (1), FGF9 (2), FGFR3 (1), GAB1 (7), HRAS (4), PIK3CA (2), STAT5A (5), TRIM24 (1), ZMYM2 (2)
57	NEGATIVE REGULATION OF FGFR SIGNALING	32	12 (0.02)	FGF17 (3), FGF2 (4), FGF6 (1), FGF8 (1), FGFR1 (1), FGFR3 (4), MAPK1 (1), MAPK3 (1), PPP2CA (1), PPP2R1A (6), SRC (1)
58	NUCLEAR SIGNALING BY ERBB4	35	14 (0.02)	APH1A (2), ERBB4 (1), ESR1 (1), GFAP (2), GH2 (1), GHR (3), NCSTN (1), PGR (1), PRLR (9), PSEN1 (1), PSEN2 (4), PSENEN (1), STAT5B (1)
59	VEGF LIGAND RECEPTOR INTERACTIONS	10	1 (0.02)	FLT4 (1), VEGFA (1)
60	SIGNALLING TO RAS	24	6 (0.02)	MAPK1 (2), MAPK11 (2), MAPK14 (1), MAPK3 (1), MAPKAPK2 (3), RALA (3)
61	SIGNALING BY FGFR1 MUTANTS	24	6 (0.02)	BCR (1), CPSF6 (1), FGF2 (3), FGF23 (1), FGF9 (2), GRB2 (1), ZMYM2 (3)
62	PRE NOTCH PROCESSING IN GOLGI	15	2 (0.02)	ATP2A2 (1), MFNG (1), NOTCH4 (2)
63	DOWNREGULATION OF ERBB2 ERBB3 SIGNALING	11	1 (0.02)	RPS27A (1), UBA52 (1)
64	CREB PHOSPHORYLATION THROUGH THE ACTIVATION OF RAS	25	5 (0.02)	ACTN2 (2), CALM1 (2), CAMK2A (2), CAMK2B (1), HRAS (1), NEFL (2)
65	PRE NOTCH TRANSCRIPTION AND TRANSLATION	18	2 (0.01)	E2F1 (1), E2F3 (1), EIF2C3 (2)
66	TGF BETA RECEPTOR SIGNALING ACTIVATES SMADS	22	3 (0.01)	FURIN (1), SMAD4 (1), TGFBR2 (1), UCHL5 (3)
67	SPRY REGULATION OF FGF SIGNALING	13	1 (0.01)	GRB2 (1), UBA52 (1)
68	ENERGY DEPENDENT REGULATION OF MTOR BY LKB1 AMPK	13	1 (0.01)	PRKAG2 (1), RHEB (1)
69	SIGNALING BY FGFR1 FUSION MUTANTS	14	1 (0.01)	CPSF6 (1), STAT5A (1)
70	INTRINSIC PATHWAY FOR APOPTOSIS	26	3 (0.01)	AKT1 (1), APAF1 (2), BCL2 (1), CASP8 (1), MAPK8 (1)

[a] Number of genes in the pathway (p).

[b] Number (proportion) of differential edges from DINGO. The proportions are divided by the number of all possible pairs of vertices ($p(p-1)/2$).

[c] Hub genes are defined by vertices that have degrees greater than 5% of the total number of differential edges.

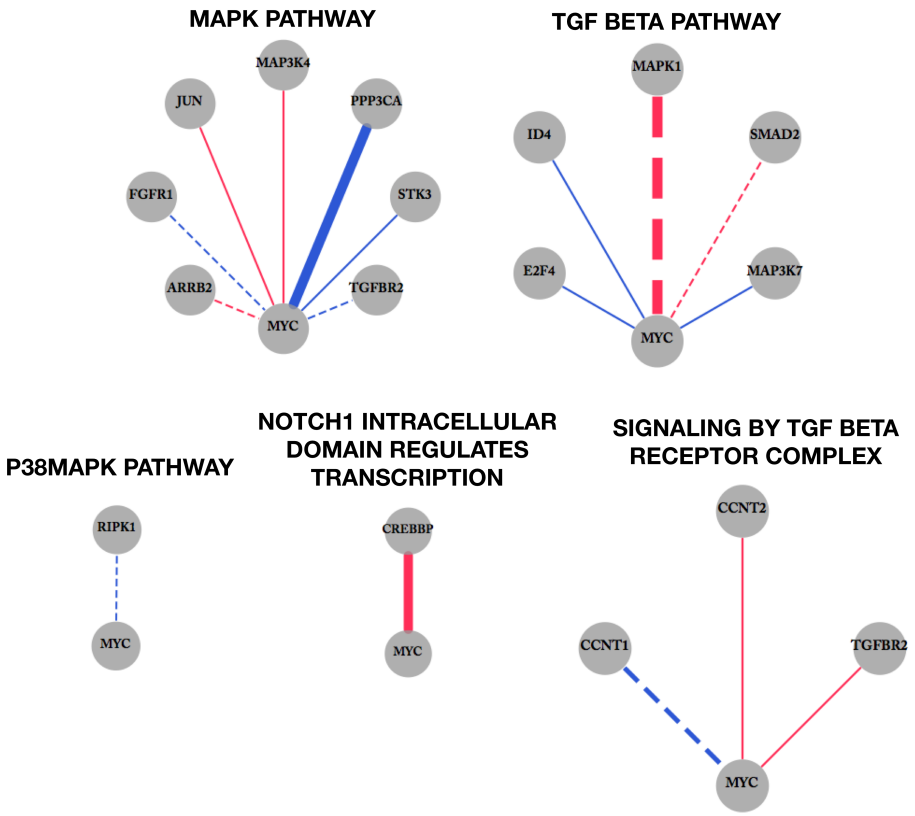

Figure S3: Differential networks for neighbors of MYC gene

Section S3 Simulation Studies

Section S3.1 Case I

The global component $\boldsymbol{\mathcal{G}}$ is generated under a specific structure. A directed acyclic graph (DAG) is a directed graph with no directed cycles. Pathways can be represented by a DAG. With a sample size (n) of 150 (75 for each group), we consider two settings, $n > p$ and $n < p$. For the $n > p$ simulation setting, we generate datasets that reflect the mRNA expression data studied in the application data example in Section 3, which includes 49 genes in the pathways. The

structure of zeros in $\boldsymbol{\mathcal{G}} = \{\mathcal{G}_{ab}\}_{49\times 49}$ is determined by the DAG of the signaling pathways in TCGA glioblastoma data (`http://cbio.mskcc.org/cancergenomics/gbm/pathways`). Specifically, we set $\mathcal{G}_{ab}$ as a random sample from `Unif`(0.2,0.8), with a randomly assigned sign when there is an edge from b to a in the pathways. Otherwise, $\mathcal{G}_{ab}$ is set to be zero. The 49 intercept terms in the first column of $\mathbf{Q}$, $\{Q_{i1} : i = 1, \ldots, p\}$ are set by random samples from `Unif`(-0.5,0.5). We consider the following 4 different simulation settings according to the effect sizes $\{Q_{i2} : i = 1, \ldots, p\}$ and noise level specified by the diagonal elements of $\boldsymbol{\Psi}$:

A1. (low effect, low noise) : $Q_{i2} \sim$ `Unif`$(0.1, 0.3)$ and $\boldsymbol{\Psi} =$ `diag`$(0.1, \ldots 0.1)$,

A2. (low effect, high noise) : $Q_{i2} \sim$ `Unif`$(0.1, 0.3)$ and $\boldsymbol{\Psi} =$ `diag`$(1, \ldots 1)$,

A3. (high effect, low noise) : $Q_{i2} \sim$ `Unif`$(0.2, 0.8)$ and $\boldsymbol{\Psi} =$ `diag`$(0.1, \ldots 0.1)$,

A4. (high effect, high noise) : $Q_{i2} \sim$ `Unif`$(0.2, 0.8)$ and $\boldsymbol{\Psi} =$ `diag`$(1, \ldots 1)$,

where all signs of $\{Q_{i2} : i = 1, \ldots, p\}$ are randomly selected.

For $n < p$ with $n = 150$ and $p = 100$ and 500, we determine the structure of $\boldsymbol{\mathcal{G}}$ by generating a scale-free network using the Barabasi-Albert algorithm (Barabási and Albert, 1999). We specify $\mathbf{Q}$ and $\boldsymbol{\Psi}$ as the setting A4 (high effect, high noise).

For the design matrix $\mathbf{X} = (\mathbf{x}_{(1)}, \ldots, \mathbf{x}_{(n)})^{\mathrm{T}}$ where $\mathbf{x}_{(i)} = (1, x_i)^{\mathrm{T}}$, the randomly selected 75 values of $\{x_i : i = 1, \ldots, n\}$ are assigned by 1 (group 1) and other values are assigned by -1 (group 2). After specifying all parameters, $\boldsymbol{\mathcal{G}}$, $\mathbf{Q}$ and $\boldsymbol{\Psi}$ and generating the design matrix $\mathbf{X}$, each row vector, $\mathbf{y}_{(i)} \in \mathbb{R}^p$ of $\mathbf{Y} = (\mathbf{y}_{(1)}, \ldots, \mathbf{y}_{(n)})^{\mathrm{T}}$ is independently generated from $N_p\left(0, \boldsymbol{\mathcal{N}}(\mathbf{x}_{(i)})\right)$ using equations (1) and (2) in Section 2.1.

Partial correlation coefficients are computed from a precision matrix by

$$\rho_{ab} = -\frac{\Omega_{ab}}{\sqrt{\Omega_{aa}\Omega_{bb}}}$$

for a precision matrix $\mathbf{\Omega} = [\Omega_{ab}]$. In our simulation study, the accuracy of the group-specific networks are evaluated by the sum of the squared error (SSE) of the partial correlation coefficients as

$$\text{SSE} = \sum_{a<b} (\hat{\rho}_{ab} - \rho_{ab})^2,$$

where $\{\hat{\rho}_{ab} : a, b \in V \text{ and } a < b\}$ are the estimates of $\{\rho_{ab} : a, b \in V \text{ and } a < b\}$. To measure the reliability of the induced ordering of the conditional dependence estimates for a network structure, the true network needs to be determined on the basis of a specific cutoff of the true conditional dependencies. We determine the

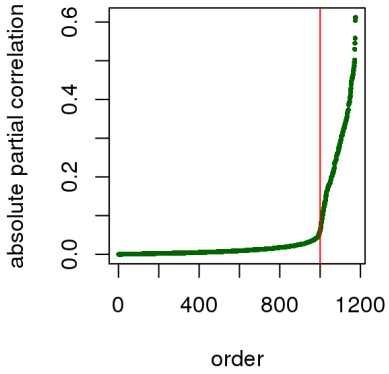

Figure S4: Orders of true absolute partial correlations versus true absolute partial correlations of the group 2 network from simulation data under the (high effect, high noise) setting. The red vertical line displays the cutoff for the true absolute partial correlations $\tau_1 = 0.06$.

cutoff (denoted by τ_1) of the absolute partial correlations for the true networks by plotting the order of the true absolute partial correlations versus the values. For example, Figure S4 displays the order of the true absolute partial correlations versus true absolute partial correlations of a group-specific network under the high effect, high noise setting. Around order 1,000 ($\tau_1 = 0.06$), the increase in the absolute values become steep and the pairs with absolute conditional dependencies greater than τ_1 are set to be edges in the true network. After setting the true network

with a cutoff, τ_1, we calculate the true positive rate (TPR), the false positive rate (FPR) and the false discovery rate (FDR) of identifying the true network structure that corresponds to a given cutoff τ_1, by varying cutoffs on the partial correlation estimates. We average those measures over 100 replicate datasets in each of the 4 scenarios. The values of the averaged 1-FPR versus TPR are displayed as average receiver operating characteristic (ROC) curves, and the values of the averaged TPR versus 1-FDR are displayed as average precision recall (PR) curves.

Section S3.2 Case II

The simulation scheme for Case I has two limitations: [1] the group-specific network structure is not obvious because the true group-specific partial correlations are not sparse; and [2] the local group-specific component is generated from a quadratic parameterization with covariates $\mathbf{x}$ in equation (1) in Section 2.1: this parameterization is the same as the assumption of our DINGO model. To overcome these limitations, we performed more realistic simulations where the group-specific data are generated from two separate GGMs that induce sparse partial correlations.

Suppose we have two groups as a covariate. We simulate the group-specific data from multivariate normal distributions with precision matrices from two GGMs where some of the edges are common to both groups and some are unique to the groups. Let x be a univariate binary covariate taking values 1 or 2. There are n_x individuals for each group. For $i \in \{1, \ldots, n_x\}$, we assume the following structural equation model for each group:

$$\mathbf{y}_{(i)} = \boldsymbol{\mathcal{G}}^x \mathbf{y}_{(i)} + \boldsymbol{\epsilon}_{(i)},$$

where $\mathbf{y}_{(i)}$ is a $p \times 1$ vector of the ith observed value of $\boldsymbol{Y}$ in group x; $\boldsymbol{\mathcal{G}}^x$ specifies the relations among variables in $V = \{1, \ldots, p\}$ in group x; and $\boldsymbol{\epsilon}_{(i)}$ is a $p \times 1$ vector following $N_p(0, \mathbf{I}_p)$ with $p \times p$ identity matrix $\mathbf{I}_p$. Note that we have the unit noise variances and zero noise covariances for both groups. To generate $\boldsymbol{\mathcal{G}}^x$ for

$x = 1$ and 2, we first construct a common structure and then add additional edges differently to the graphs for the two groups. As the common graph, we consider the random DAG model (Erdős and Rényi, 1960), which is called the ER model. The ER model constructs a DAG of p vertices by connecting vertices randomly. Each edge is included in the graph with probability α independent from all other edges. There are $(p-1)\alpha$ numbers of edges in expectation and α controls the sparsity in the common structure. The $\boldsymbol{\mathcal{G}}^x$ for $x = 1$ and 2 are generated as follows: (1) let $\boldsymbol{\mathcal{G}}$ be a $p \times p$ zero matrix; (2) replace each entry of $\boldsymbol{\mathcal{G}}$ with a uniform random sample from the $[0.5, 1]$ interval with probability α (the nonzero entries are selected with the restriction of no directed cycle); (3) set $\boldsymbol{\mathcal{G}}^1 = \boldsymbol{\mathcal{G}}^2 = \boldsymbol{\mathcal{G}}$; (4) replace a randomly selected entry from the zero entries of $\boldsymbol{\mathcal{G}}^1$ and $\boldsymbol{\mathcal{G}}^2$ with a uniform random sample from the $[0.5, 1]$ interval. With the restriction of no directed cycle, step (4) is repeated $\gamma \times K$, where K is the number of nonzero elements in $\boldsymbol{\mathcal{G}}$ and γ controls the ratio of the number of individual edges to the number of common edges (heterogeneity of the graphs). Then, $\mathbf{y}_{(i)} \sim N_p(\mathbf{0}, (\mathbf{I}-\boldsymbol{\mathcal{G}}^1)^{-1}(\mathbf{I}-\boldsymbol{\mathcal{G}}^1)^{-\mathrm{T}})$ for the individual i in group 1 and $\mathbf{y}_{(i)} \sim N_p(\mathbf{0}, (\mathbf{I}-\boldsymbol{\mathcal{G}}^2)^{-1}(\mathbf{I}-\boldsymbol{\mathcal{G}}^2)^{-\mathrm{T}})$ for the individual i in group 2. We can generate the n_1 and n_2 samples separately from the distributions. We consider 4 simulation settings in Table S5, with the number of vertices p, the level of sparsity α, and the level of heterogeneity γ.

Table S5: Simulation setting

no. of vertices	level of sparsity	level of heterogeneity	no. of edges (proportion)	
p	α	γ	group 1	group 2
50	0.01	0.25	40 (0.033)	42 (0.034)
50	0.01	0.75	81 (0.066)	56 (0.046)
200	0.005	0.25	530 (0.027)	224 (0.011)
200	0.005	0.75	1901 (0.096)	1368 (0.069)

This simulation setting gives sparsity values from 90% to 99% to the group-specific GGMs. Figure S5 displays the simulated group-specific GGMs, induced by $\boldsymbol{\mathcal{G}}^1$ and $\boldsymbol{\mathcal{G}}^2$ under the scenario $(p, \alpha, \gamma) = (50, 0.01, 0.75)$. There are 34 common edges, and 47 and 22 unique edges to the group 1 and group 2 GGMs. We set the samples sizes $n_1 = n_2 = 75$.

We compared our DINGO method with the two separate estimations, MLE and GLasso. DINGO and MLE provide saturated (non-sparse) group-specific partial correlations. For the DINGO and MLE, we change the cutoffs for the estimated

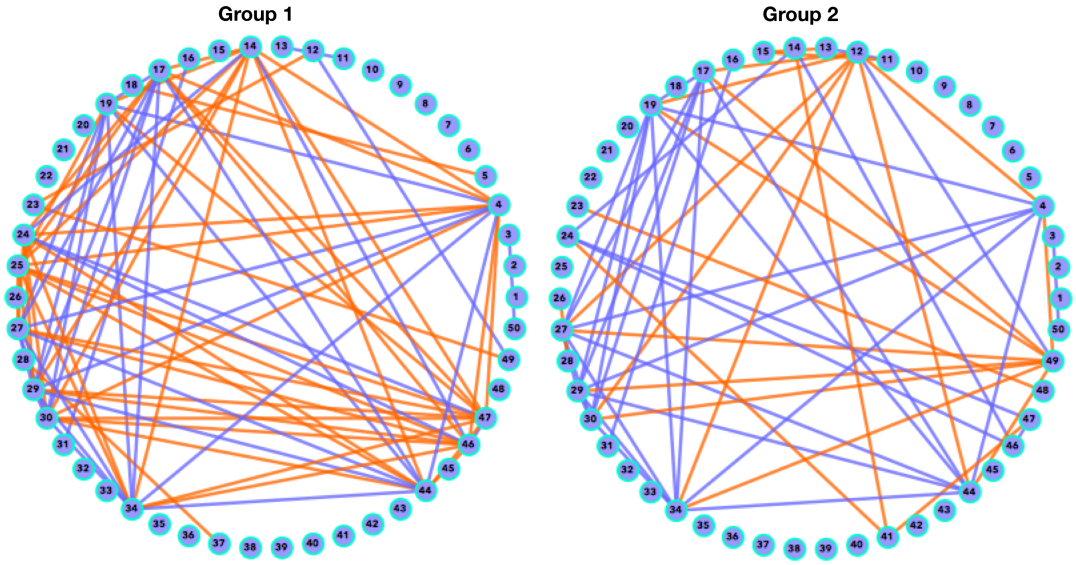

Figure S5: The simulated group-specific GGMs induced by $\boldsymbol{\mathcal{G}}^1$ and $\boldsymbol{\mathcal{G}}^2$ under the setting $(p, \alpha, \gamma) = (50, 0.01, 0.75)$. There are 34 common edges (purple), and 47 and 22 unique edges (orange) to the group 1 and group 2.

partial correlations to draw the receiver operating characteristic (ROC) and precision recall (PR) curves. For the GLasso, we changed the tuning parameter η in equation (3) of the main text from 0.001 to 1 for the two $p = 50$ settings and from 0.01 to 2 for the two $p = 200$ settings to obtain the ROC and PR curves. For SSE, we selected the tuning parameter η, using the Bayesian information criterion (BIC) for

the GLasso. For MLE and DINGO, we used the saturated group-specific partial correlations for the MLE and DINGO. Since MLE is not valid for the sample size of 75 and the number of variables of 200, we display results for the GLasso and DINGO for the two $p = 200$ settings. The simulation results are displayed in Figure S6 - Figure S9 based on 100 replications of the data. Our DINGO model performs better in terms of SSE and ROC curves under all four settings. Although from the PR curves, we see that GLasso performs better than DINGO when the recall is small, DINGO generally performs better for almost all regions of the recall.

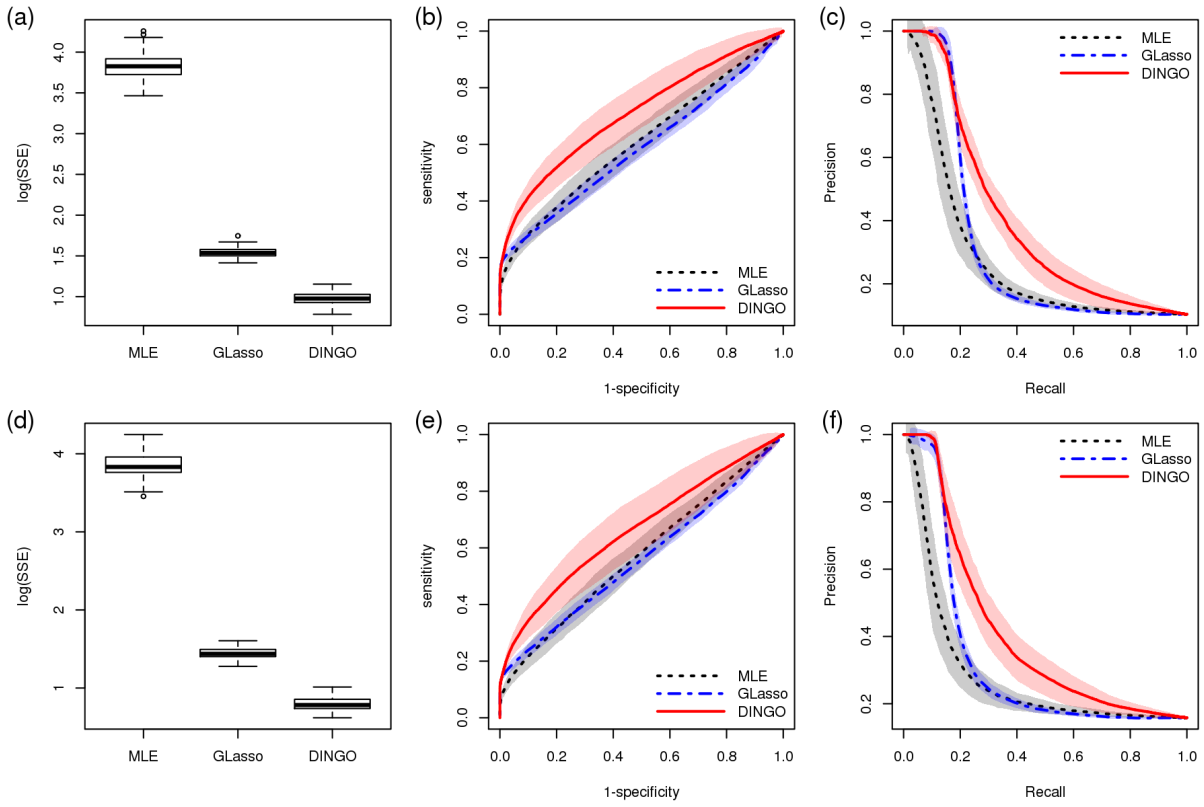

Figure S6: Simulation results for the $(p, \alpha, \gamma) = (50, 0.01, 0.25)$ scenario from 100 simulation datasets. Group 1 network: (a) boxplots of log(SSE); (b) receiver operating characteristic (ROC) curves; (c) precision recall (PR) curves.; Group 2 network: (d) boxplots of log(SSE); (e) ROC curves; (f) PR curves. Regions of one standard error for the y-axis are shaded with the corresponding colors of the ROC and PR curves.

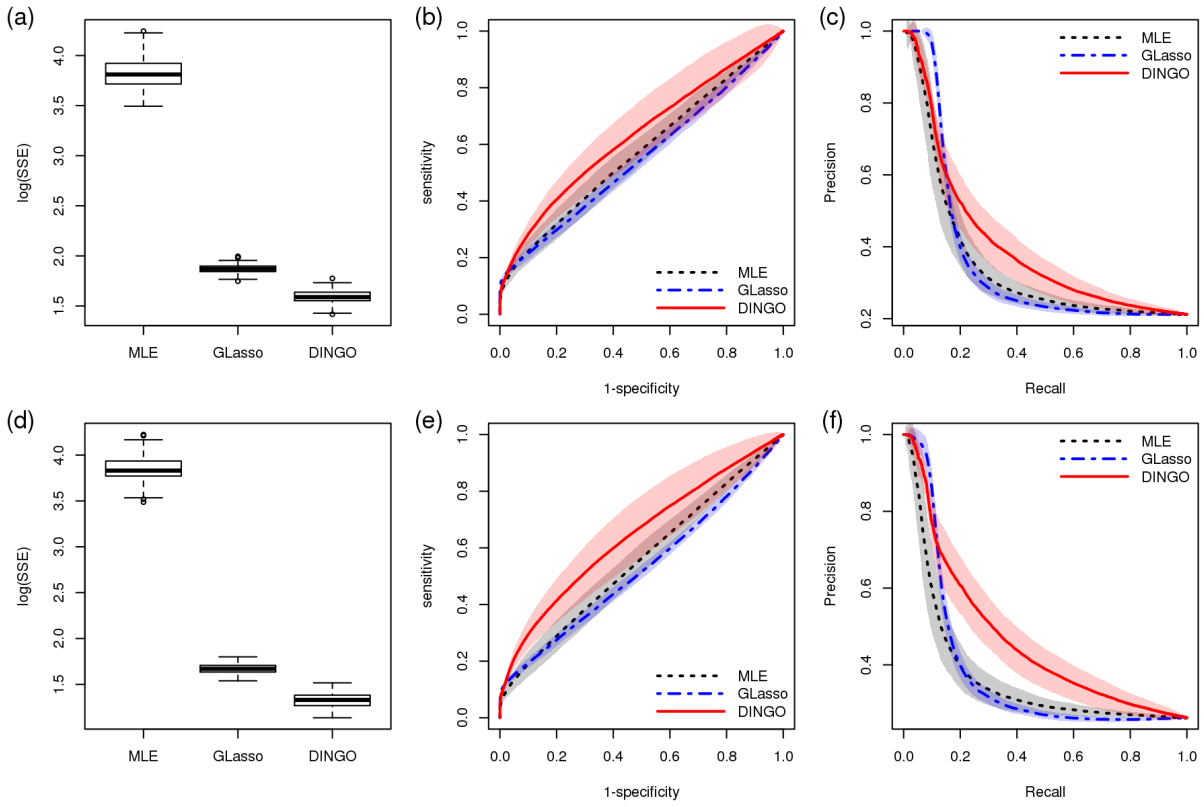

Figure S7: Simulation results for the $(p, \alpha, \gamma) = (50, 0.01, 0.75)$ scenario from 100 simulation datasets. Group 1 network: (a) boxplots of log(SSE); (b) receiver operating characteristic (ROC) curves; (c) precision recall (PR) curves.; Group 2 network: (d) boxplots of log(SSE); (e) ROC curves; (f) PR curves. Regions of one standard error for the y-axis are shaded with the corresponding colors of the ROC and PR curves.

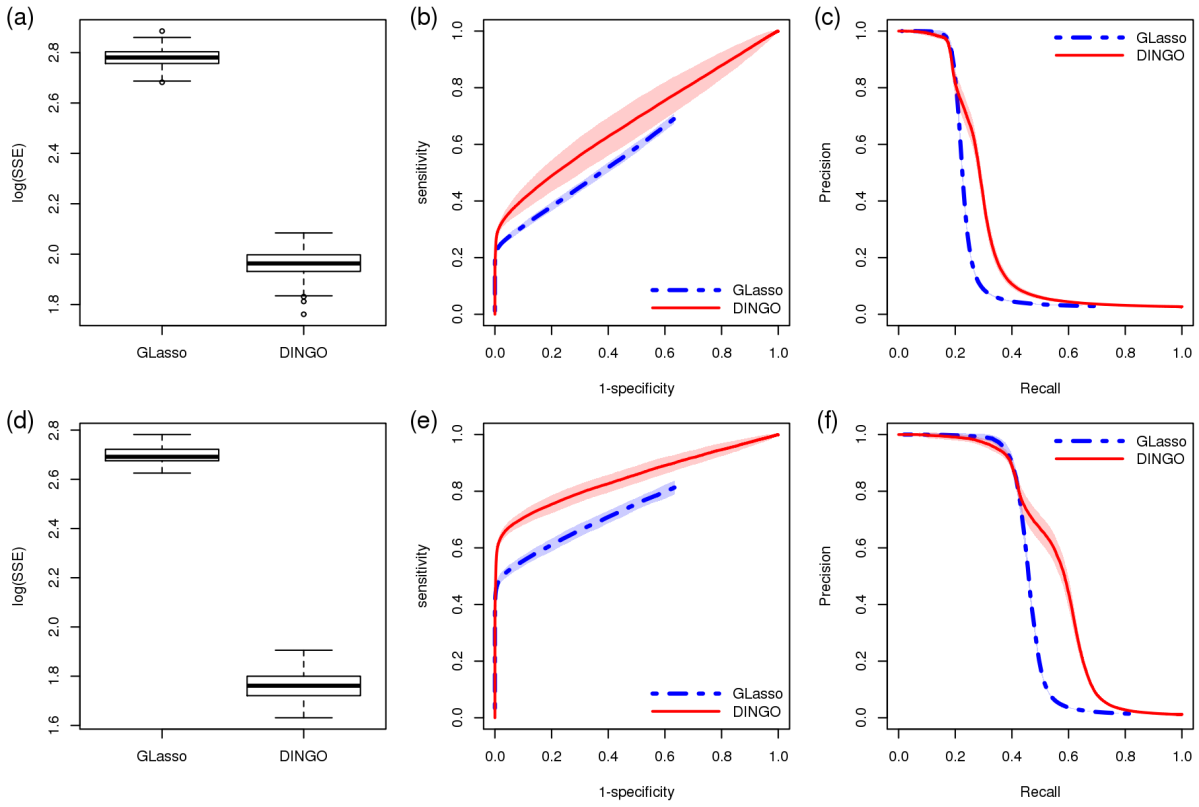

Figure S8: Simulation results for the $(p, \alpha, \gamma) = (200, 0.005, 0.25)$ scenario from 100 simulation datasets. Group 1 network: (a) boxplots of log(SSE); (b) receiver operating characteristic (ROC) curves; (c) precision recall (PR) curves.; Group 2 network: (d) boxplots of log(SSE); (e) ROC curves; (f) PR curves. Regions of one standard error for the y-axis are shaded with the corresponding colors of the ROC and PR curves.

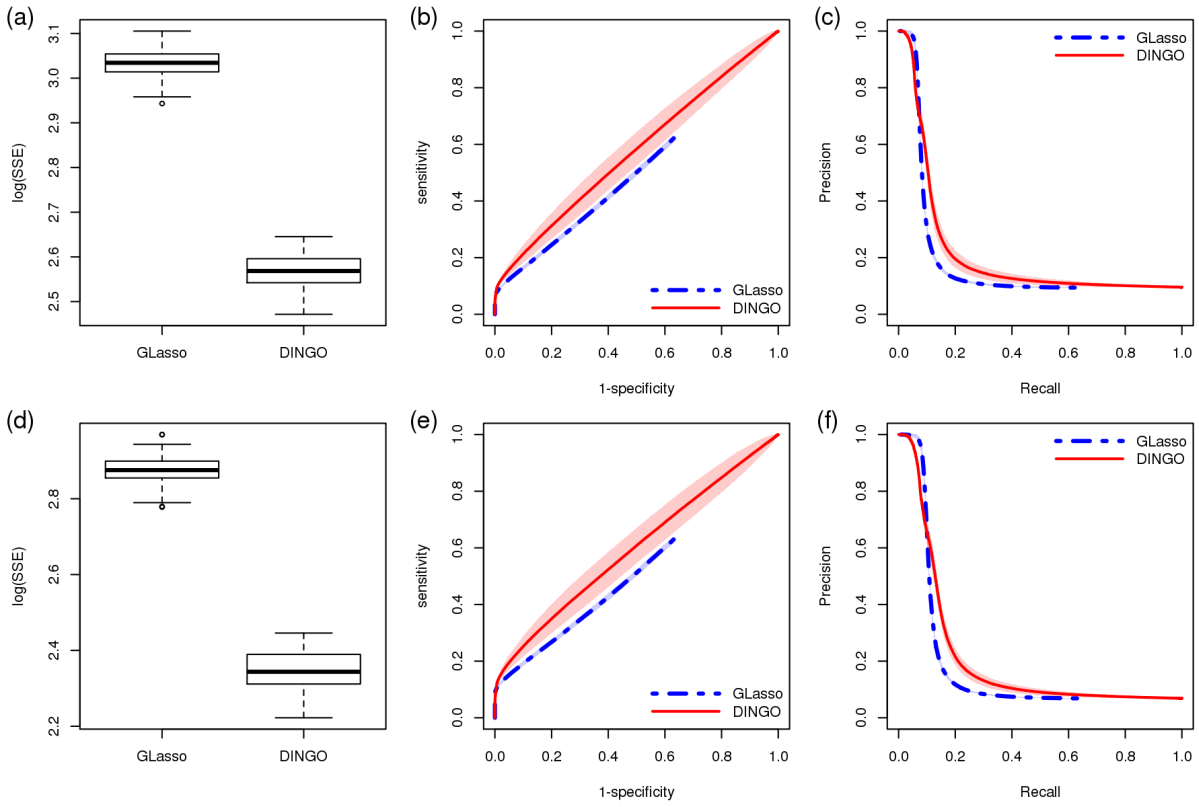

Figure S9: Simulation results for the $(p, \alpha, \gamma) = (200, 0.005, 0.75)$ scenario from 100 simulation datasets. Group 1 network: (a) boxplots of log(SSE); (b) receiver operating characteristic (ROC) curves; (c) precision recall (PR) curves.; Group 2 network: (d) boxplots of log(SSE); (e) ROC curves; (f) PR curves. Regions of one standard error for the y-axis are shaded with the corresponding colors of the ROC and PR curves.

Section S4 Comparison of DINGO on real data

We compare DINGO to a much larger set of methods on real data. Using the mRNA expression data, used in Section 3, for genes involved in the RTK/PI3K, p53, and Rb signaling pathways, we explicitly compared our group-specific and differential network methods with other methods in the literature.

Group-specific networks: We first compare the estimated local group-specific networks obtained from DINGO with those obtained from the following four methods:

(1) **Maximum likelihood estimation (MLE)** estimates the precision matrix by maximizing the likelihood.

(2) **Graphical Lasso (GLasso)** (Friedman et al., 2008) estimates a sparse precision matrix by maximizing the L1 penalized likelihood.

(3) **Weighted correlation network analysis (WGCNA)** (Langfelder and Horvath, 2008) estimates the weighted correlation network using marginal correlation.

(4) **GeneNet** (Schäfer and Strimmer, 2005) provides shrinkage estimates of partial correlations.

All the above methods provide weighted edges for assessing the network topology. After sorting the absolute values of the weights in decreasing order, we took the top 120 (10% of the total number of edges) for each method for performance comparisons. We then calculated the number of common edges for all pairs of methods. The results are shown in Table S6 for long-term survivors (LTSs) and Table S7 for short-term survivors (STSs), respectively. We observed some degree of similarity between the methods and, in particular, found that our DINGO method is closer to GeneNet and GLasso. We conjecture this is due to the use of penalized partial correlation estimates by these three methods.

Table S6: Number (%) of common edges for LTSs

	DINGO	MLE	GLasso	WGCNA	GeneNet
DINGO	-	30 (25)	52 (43.33)	48 (40)	60 (50)
MLE	-	-	31 (25.83)	21 (17.5)	56 (46.67)
GLasso	-	-	-	55 (45.83)	55 (45.83)
WGCNA	-	-	-	-	35 (29.17)
GeneNet	-	-	-	-	-

Table S7: Number (%) of common edges for STSs

	DINGO	MLE	GLasso	WGCNA	GeneNet
DINGO	-	30 (25)	57 (47.5)	55 (45.83)	56 (46.67)
MLE	-	-	38 (31.67)	27 (22.5)	56 (46.67)
GLasso	-	-	-	56 (46.67)	60 (50)
WGCNA	-	-	-	-	48 (40)
GeneNet	-	-	-	-	-

Differential network: We also compare the differential networks from DINGO with those from a differential coexpression analysis method using DCGL (Differential Coexpression Analysis and Differential Regulation Analysis of Gene Expression Microarray Data) R package (Liu et al., 2010). The DCGL is a tool for identifying differentially coexpressed genes and links based on marginal correlation. Applying DCGL to our dataset discovered 71 edges. For our DINGO method, we use the differential score from a bootstrap procedure and assess the degree of similarity. For varying cutoffs on the differential score, we estimate the similarity of the two networks using the F-score (Knaack et al., 2014), which is calculated as follows. The number of edges that are present in both networks is defined as *common edges* and the *precision* and *recall* are defined as the ratio of the number of common edges to the number of edges for both the DINGO and DGCL networks. The F-score is then defined as the harmonic mean of precision and recall, and reflects the degree of similarity.

Figure S10 displays the number of edges detected by DINGO, varying cutoffs on the differential score versus F-score to compare the networks from DINGO and DCGL. We observed that the degree of similarity shows an increasing trend, i.e.,

both methods have a few edges in common. However, the overall degree of similarity is low, which we conjecture is due to different estimation techniques in the respective algorithms. DCGL uses a filtering technique before fitting the differential network and handles marginal correlations, i.e., two genes at a time. In constrast, DINGO uses a model-based approach to fit group-specific partial correlations, which obtains more refined associations than marginal correlations because they look at the pathway/gene-set as a whole.

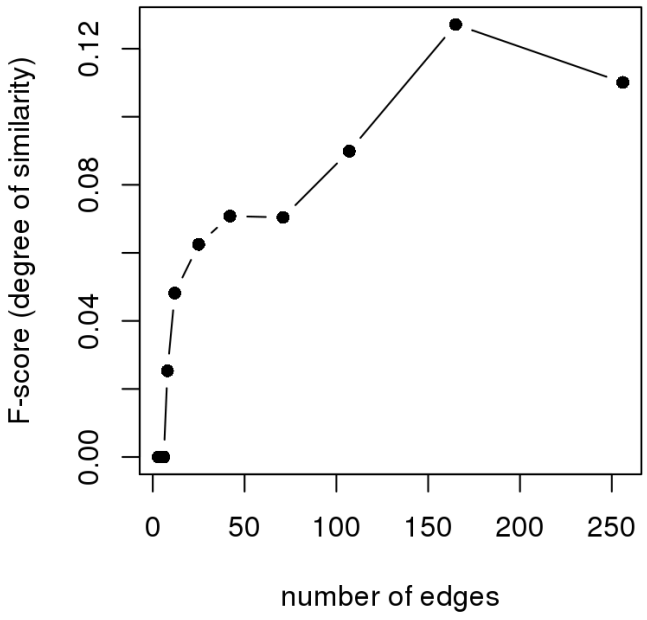

Figure S10: Number of edges versus F-score to compare networks obtained from DINGO and DCGL.

Section S5 Supplementary figures and tables

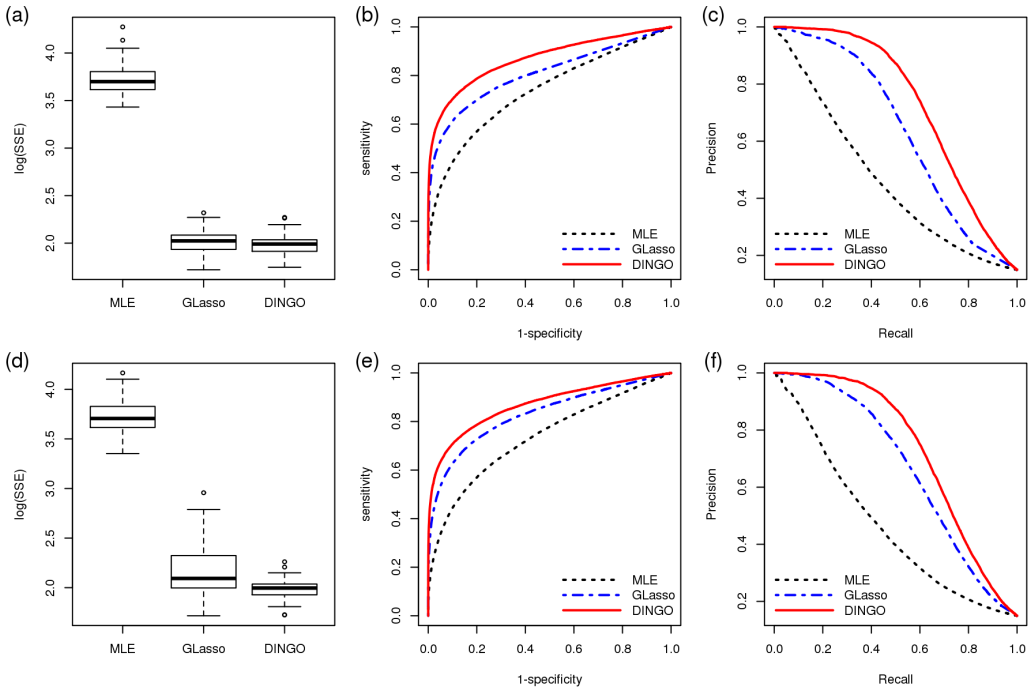

Figure S11: Simulation results for the (low effect, low noise) scenario from 100 simulation datasets. Group 1 network: (a) boxplots of log(SSE); (b) receiver operating characteristic (ROC) curves; (c) precision recall (PR) curves.; Group 2 network: (d) boxplots of log(SSE); (e) ROC curves; (f) PR curves.

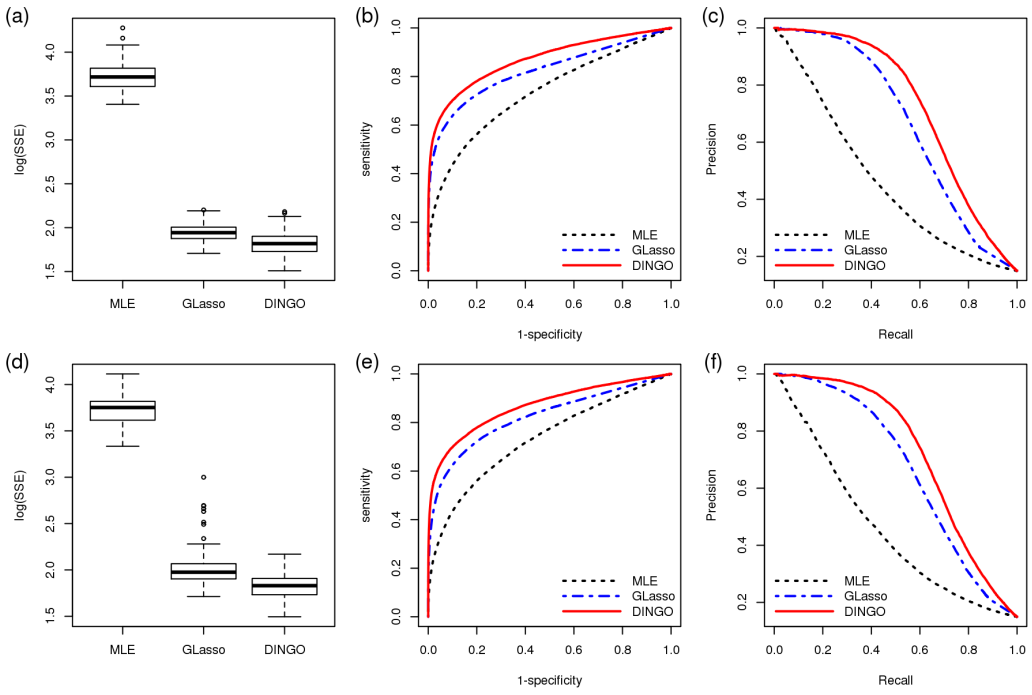

Figure S12: Simulation results for the (low effect, high noise) scenario from 100 simulation datasets. Group 1 network: (a) boxplots of log(SSE); (b) receiver operating characteristic (ROC) curves; (c) precision recall (PR) curves.; Group 2 network: (d) boxplots of log(SSE); (e) ROC curves; (f) PR curves.

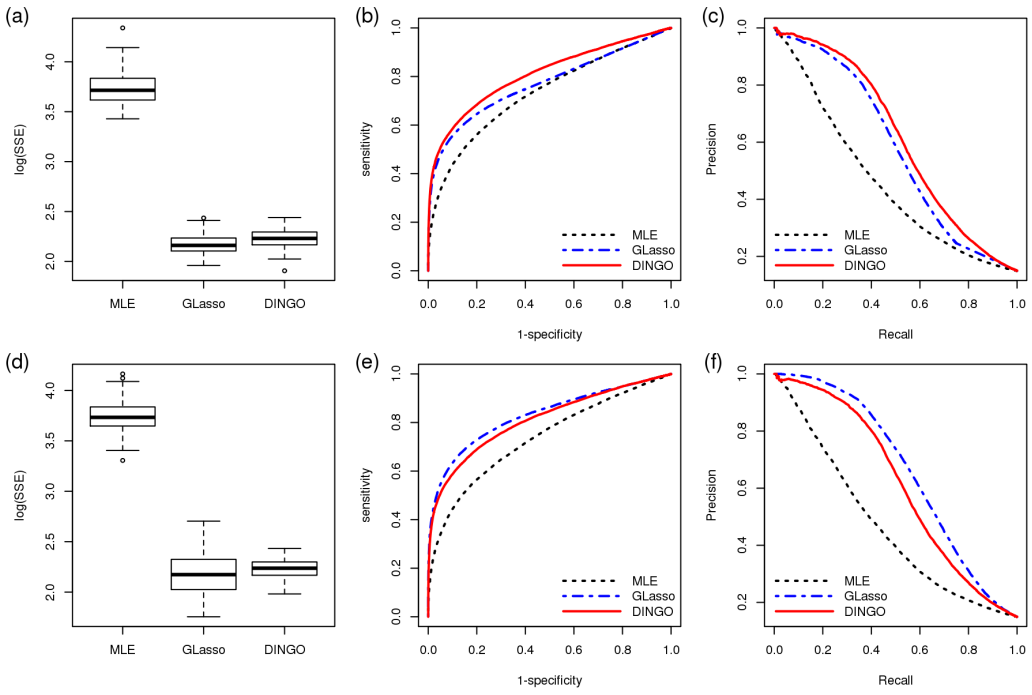

Figure S13: Simulation results for the (high effect, low noise) scenario from 100 simulation datasets. Group 1 network: (a) boxplots of log(SSE); (b) receiver operating characteristic (ROC) curves; (c) precision recall (PR) curves.; Group 2 network: (d) boxplots of log(SSE); (e) ROC curves; (f) PR curves.

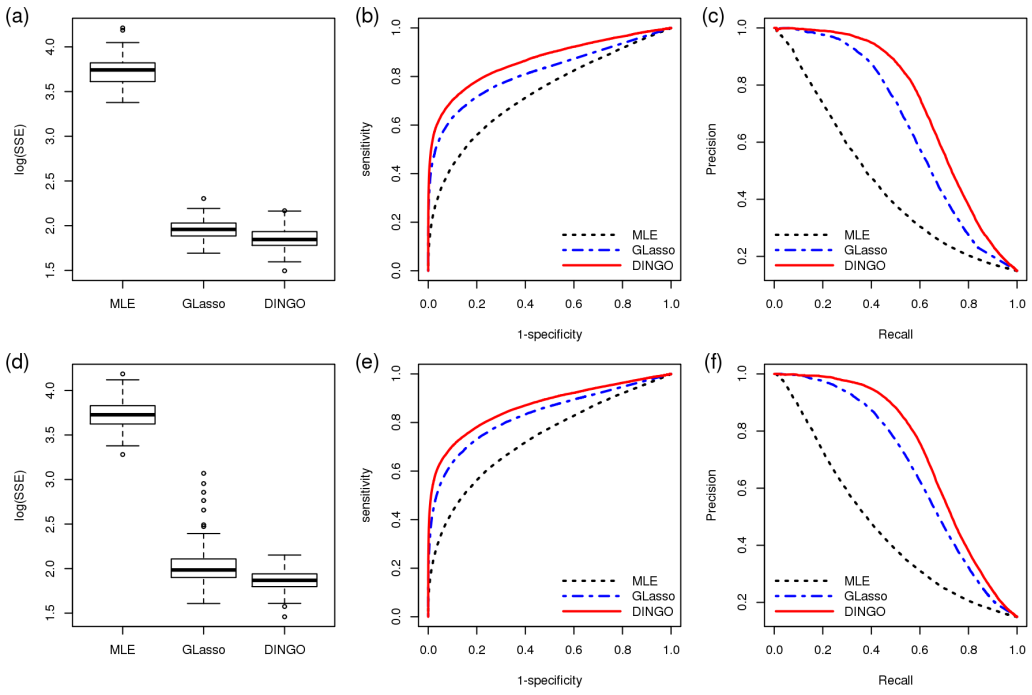

Figure S14: Simulation results for the (high effect, high noise) scenario from 100 simulation datasets. Group 1 network: (a) boxplots of log(SSE); (b) receiver operating characteristic (ROC) curves; (c) precision recall (PR) curves.; Group 2 network: (d) boxplots of log(SSE); (e) ROC curves; (f) PR curves.

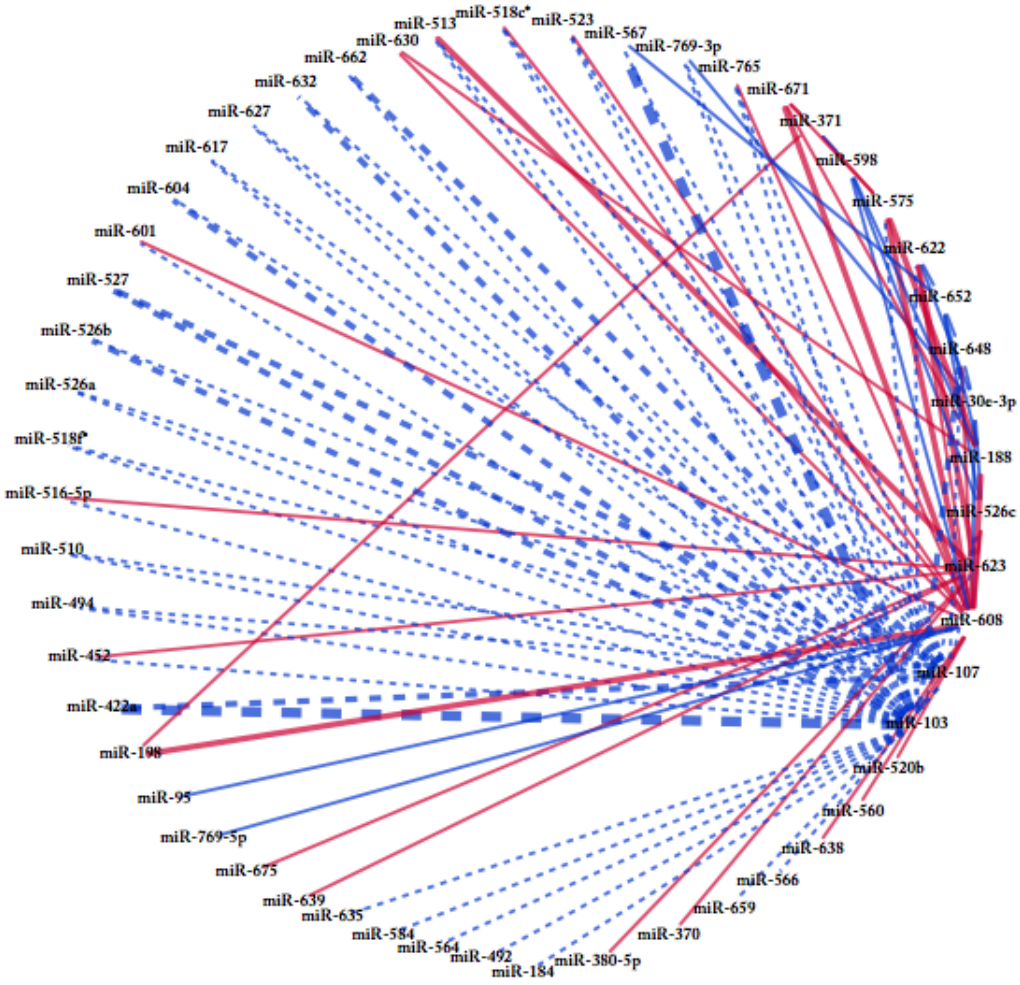

Figure S15: Differential networks between the LTSs and STSs of glioblastoma estimated from microRNA expression. The vertices are ordered by degrees (number of connections). The blue (red) edges indicate a positive (negative) score. The solid (dashed) lines represent conserved (differential) signs in the dependencies between the long-term survivors and short-term survivors. The thickness of the edges corresponds to the strength of the associations, with stronger scores having greater thickness.

Table S8: Notations

Notation	Type	Definition
n	scalar	sample size
p	scalar	number of genes
$\mathbf{y}$	$p \times 1$ vector	observations for p genes over a sample
x	scalar	binary covariate
$\mathbf{Y}$	$n \times p$ matrix	data matrix for $\mathbf{y}$ for all n samples
$\mathbf{X}$	$n \times q$ matrix	design matrix for covariates for all n samples
$\mathbf{x}$	$q \times 1$ vector	row of the design matrix $\mathbf{X}$
V	set	set of vertices
E	set	set of edges
$\boldsymbol{\mathcal{N}}$	$p \times p$ matrix	GGM (precision matrix) of $\mathbf{y}$
$\boldsymbol{\mathcal{N}}(x)$	$p \times p$ matrix function	group-specific GGMs (precision matrices) of $\mathbf{y}$
$\boldsymbol{\mathcal{G}}$	$p \times p$ matrix	global component (coefficient matrix for the global network model)
$\boldsymbol{\epsilon}$	$p \times 1$ vector	residual vector after taking out effects of the global relations
$\boldsymbol{\mathcal{L}}$	$p \times p$ matrix	local GGM (precision matrix of $\boldsymbol{\epsilon}$)
$\boldsymbol{\mathcal{L}}(x)$	$p \times p$ matrix function	local group-specific component
$\mathbf{Q}$	$p \times q$ matrix	coefficient parameter in the precision regression model
$\boldsymbol{\Psi}$	$p \times p$ diagonal matrix	variance parameter in the precision regression model
$\rho_{ab}^{(i)}$ $(\hat{\rho}_{ab}^{(i)})$	scalar	partial correlation (estimate) between vertices a and b for ith group
$\phi_{ab}^{(i)}$ $(\hat{\phi}_{ab}^{(i)})$	scalar	Fisher's Z transformation of $\rho_{ab}^{(i)}$ $(\hat{\rho}_{ab}^{(i)})$
s_{ab}^{B}	scalar	bootstrap estimate of standard error for the difference, $\hat{\phi}_{ab}^{(1)} - \hat{\phi}_{ab}^{(2)}$
$\delta_{ab}^{(12)}$	scalar	differential score for the edge, a-b, between group 1 and group 2

Section S6 R code to use DINGO package

The DINGO package can be downloaded at `http://odin.mdacc.tmc.edu/~vbaladan/Veera_Home_Page/Software.html`.

```
> library("DINGO")
```

The example data are loaded as follows.

```
> data(gbm)
> dim(gbm)
[1] 156  19
> gbm[1:5,1:5]
               x       AKT1        AKT2       AKT3     FOXO1A
TCGA-02-2466   1 -1.0029131 -2.10830563  0.1090082  0.6075851
TCGA-02-2483   1  1.2592399  0.15877579  0.1911298 -0.2751667
TCGA-02-2485   1  1.0144639 -0.03734893  8.0576114  1.5937215
TCGA-02-2486   1  0.4817741 -0.97431724 -0.5723904 -0.3706160
TCGA-06-0122  -1  0.4005927  1.08527185  1.3512510 -0.6834011
> x = gbm[,1]
> expDat = gbm[,2:19]
```

The example dataset includes the group covariate, 1 for LTSs and -1 for STSs (in the first column from the left) and standardized mRNA expression data for 18 genes included in the PI3K signaling pathway for 156 TCGA GBM patients (2nd-19th columns). We have 73 patients in the STSs and 83 patients in the LTSs.

```
> table(x)
x
-1  1
73 83.
```

We fit our DINGO model and calculate the differential scores from 10 bootstrap samples as follows:

```
> fit = dingo(dat=expDat,x=x,diff.score=T,B=10)
> names(fit)
 [1] "genepair"   "levels.x"   "R1"         "R2"         "boot.diff"
 [6] "diff.score" "rho"        "P"          "Q"          "Psi" .
```

All possible pairs of the genes are listed in

```
> head(fit$genepair)
  gene1  gene2
1  AKT1   AKT2
2  AKT1   AKT3
3  AKT2   AKT3
4  AKT1 FOXO1A
5  AKT2 FOXO1A
6  AKT3 FOXO1A.
```

The coding scheme for the design matrix $\mathbf{X}$ is displayed in

```
> fit$levels.x
[1]  1 -1.
```

This means that the 2-dimensional covariate vector is $\mathbf{x} = \begin{pmatrix} 1 & 1 \end{pmatrix}^{\mathrm{T}}$ for LTSs and $\mathbf{x} = \begin{pmatrix} 1 & -1 \end{pmatrix}^{\mathrm{T}}$ for STSs. We display the group-specific partial correlations for LTSs and STSs as

```
> fit$R1[1:3]
[1]  0.02178679 -0.01730649 -0.01263468
> fit$R2[1:3]
[1]  0.014096913 -0.006587251  0.005055453.
```

The order of those vectors is the same as the order of the rows of "genepair". The differences of the Fisher's Z transformed partial correlations between LTSs and STSs from $B = 10$ bootstrap samples are displayed as

```
> head(fit$boot.diff)
             [,1]          [,2]         [,3]         [,4]          [,5]
[1,] -0.008095735  0.0030048864 -0.004958005  0.001727282 -0.0147764719
[2,]  0.010092920  0.0037487056 -0.008436010 -0.160430055  0.0007262764
[3,] -0.004409726 -0.0255130082 -0.013268548  0.010232124  0.0104620936
[4,]  0.098827078  0.0009669561  0.037737915 -0.004053196  0.0740319665
[5,]  0.014394249  0.0141214544  0.021714127  0.002786202 -0.0908931983
[6,] -0.017929906 -0.0107630001  0.025720164 -0.065634774 -0.0262401084
             [,6]          [,7]          [,8]         [,9]         [,10]
[1,]  0.011720537  0.0274612409 -0.0012454039  0.001225594 -0.012686608
[2,] -0.054668542 -0.0045957079 -0.0096877769 -0.001578630 -0.011578423
[3,] -0.026316036 -0.0192654456 -0.0011782967  0.031197794 -0.005915831
[4,]  0.016206441  0.0019505684 -0.0358101174 -0.028695654  0.024097544
[5,]  0.005713372  0.0074173716 -0.0006686498  0.099036700  0.016786118
[6,]  0.034691583  0.0006974551 -0.0086358596 -0.112107824  0.016581086.
```

The order of the rows corresponds to the order of the "genepair" and the columns correspond to the bootstrap samples. Differential scores for all edges that correspond to the order of "genepair" are

```
> fit$diff.score[1:3]
[1] 0.6223829 -0.2094640 -0.9787110.
```

The differential edges can be listed as

```
> fit$genepair[abs(fit$diff.score)>2,]
     gene1  gene2
```

```
62    GAB1 PIK3CA
64   PDPK1 PIK3CA
92    AKT1 PIK3CG
93    AKT2 PIK3CG
105 PIK3CD PIK3CG
119 PIK3CD PIK3R1
134 PIK3CD PIK3R2
148 PIK3CA   PTEN
150 PIK3CD   PTEN
151 PIK3CG   PTEN.
```

Section S7 Computation times for pathway-based DINGO analysis

While pathways used in our analysis (Section 3) have been implicated in GBM in prior studies, the exact nature of the pathway components has not been studied with respect to differential patterns of activation/inactivation related to the patient prognostic groups. Our re-analysis focuses on the exact pathway breakages using data from multiple platforms, which sheds a completely different light on the various biological processes involved in GBM progression.

Although our DINGO model can theoretically handle genome-wide data under the two-group scenarios, we use a pathway-based approach for illustration due to the limited sample size (n=156) of the TCGA GBM dataset as well as the computational times involved in model fitting. In step 1, we use graphical Lasso (GLasso) to estimate the global component. GLasso is designed for high-dimensional data and relies on the assumption that the concentration matrix is sparse (most genes are conditionally independent under a normality assumption). A consistent estimation of the network is proved under a set of assumptions, including high dimensionality,

$p \gg n$ (Meinshausen and Bühlmann, 2006). However the algorithm can efficiently handle data involving $\sim 1,000$ genes and hundreds of samples (Peng et al., 2009). In step 2, we estimate the local group-specific component using the precision regression model, which includes $3p$ parameters for the two-group setting, where p is the number of genes. Because we are not explicitly exploiting a sparsity assumption on the parameters, estimating all $3p$ parameters is untenable for a genome-wide application involving more than 22K genes.

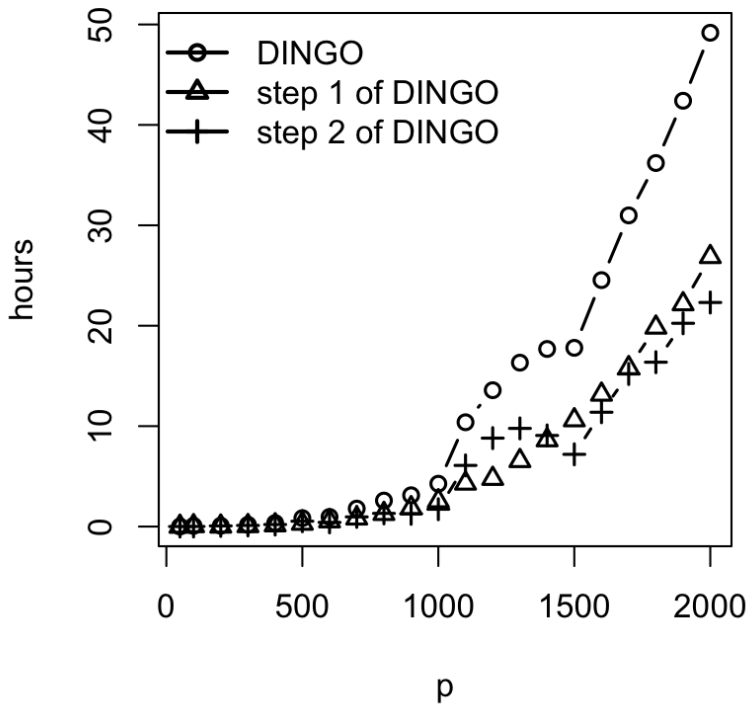

Figure S16: Computation times for step 1 and step 2 of DINGO algorithm

Figure S16 displays the computation times for step 1 and step 2 of DINGO when the sample size is equal to 156 (the same as our application data) using a Linux server with a 2.67 GHz Intel processor and 96GB of RAM. For step 1 of DINGO,

we estimate the global component using GLasso, with the tuning parameter selected among 100 candidates. For step 2 of DINGO, we estimate the local group-specific component using the EM algorithm. As p increases, the computation time increases exponentially. DINGO took about 4 hours for $p = 1000$ and took about 2 days for $p = 2000$. Combined with step 3 of DINGO to calculate differential scores from the bootstrap procedure, the DINGO algorithm might take more than 2 days for $p > 2000$. With our current implementation, handling genome-wide data with p more than 22K is infeasible; we will explore faster implementation in the future using graphical processing units (GPUs).

As an alternative to fitting genome-wide data, one can pre-filter genes before using our DINGO model. Specifically, a gene is kept if the gene has high marginal correlations with many other genes. However, as we emphasized throughout the paper, a pathway-based approach allows for refined biological interpretations, especially for practitioners who tend to think in terms of the pathway-based disruptions involved in disease progression for potential downstream translational use.

For the GBM case study, Figure S17 displays histograms for the number of genes per pathway from the three well-established databases: KEGG, BIOCARTA, and REACTOME. The medians (99% quantiles) from all three databases were less than 50 (400). Our illustrative examples are based on these numbers, i.e., $50 - 600$ genes can cover almost all *known* pathways. For $p = 50$ and $p = 600$, step 1 and step 2 of DINGO take around 20 seconds and 57 minutes, respectively, which makes it feasible to conduct pathway-based differential network analysis.

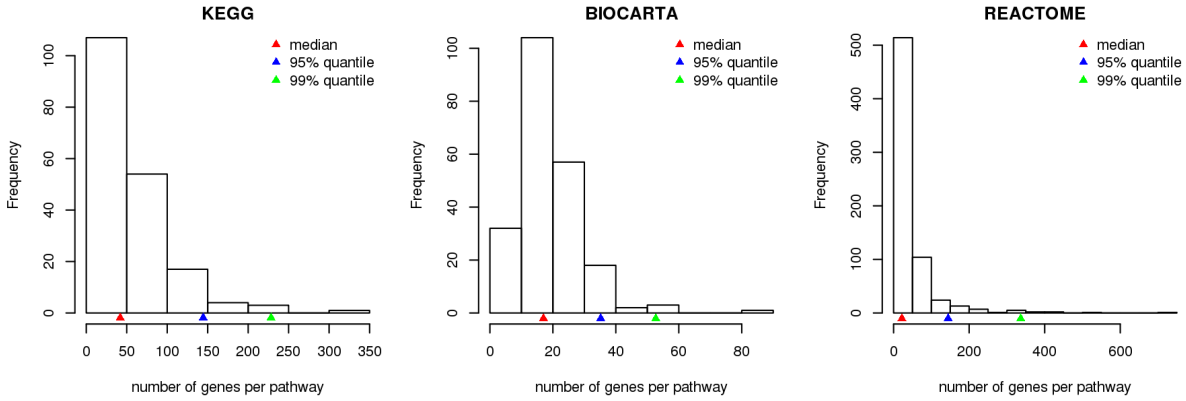

Figure S17: Histograms of number of genes per pathway in KEGG, BIOCARTA, and REACTOME.

References

Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *science*, 286(5439):509–512.

Bello, L., Francolini, M., Marthyn, P., Zhang, J., Carroll, R. S., Nikas, D. C., Strasser, J. F., Villani, R., Cheresh, D. A., and Black, P. M. (2001). αvβ3 and αvβ5 integrin expression in glioma periphery. *Neurosurgery*, 49(2):380–390.

Desgrosellier, J. S. and Cheresh, D. A. (2010). Integrins in cancer: biological implications and therapeutic opportunities. *Nature Reviews Cancer*, 10(1):9–22.

Dienstmann, R., Rodon, J., Prat, A., Perez-Garcia, J., Adamo, B., Felip, E., Cortes, J., Iafrate, A., Nuciforo, P., and Tabernero, J. (2014). Genomic aberrations in the fgfr pathway: opportunities for targeted therapies in solid tumors. *Annals of oncology*, 25(3):552–563.

Erdős, P. and Rényi, A. (1960). On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 5:17–61.

Fan, X., Khaki, L., Zhu, T. S., Soules, M. E., Talsma, C. E., Gul, N., Koh, C., Zhang, J., Li, Y.-M., Maciaczyk, J., et al. (2010). Notch pathway blockade depletes cd133-positive glioblastoma cells and inhibits growth of tumor neurospheres and xenografts. *Stem cells*, 28(1):5–16.

Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441.

Guo, W. and Giancotti, F. G. (2004). Integrin signalling during tumour progression. *Nature reviews Molecular cell biology*, 5(10):816–826.

Hanahan, D. and Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell*, 144(5):646–674.

Hoff, P. D. and Niu, X. (2012). A covariance regression model. *Statistica Sinica*, 22:729–753.

Knaack, S. A., Siahpirani, A. F., and Roy, S. (2014). A pan-cancer modular regulatory network analysis to identify common and cancer-specific network components. *Cancer Inform*, 13(Suppl 5):69–84.

Langfelder, P. and Horvath, S. (2008). Wgcna: an r package for weighted correlation network analysis. *BMC bioinformatics*, 9(1):559.

Liu, B.-H., Yu, H., Tu, K., Li, C., Li, Y.-X., and Li, Y.-Y. (2010). Dcgl: an r package for identifying differentially coexpressed genes and links from gene expression microarray data. *Bioinformatics*, 26(20):2637–2638.

McLendon, R., Friedman, A., Bigner, D., Van Meir, E. G., Brat, D. J., Mastrogianakis, G. M., Olson, J. J., Mikkelsen, T., Lehman, N., Aldape, K., et al. (2008). Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*, 455(7216):1061–1068.

Meinshausen, N. and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics*, pages 1436–1462.

Miller, K. S. (1981). On the inverse of the sum of matrices. *Mathematics Magazine*, 54(2):67–72.

Nakada, M., Kita, D., Watanabe, T., Hayashi, Y., Teng, L., Pyko, I. V., and Hamada, J.-I. (2011). Aberrant signaling pathways in glioma. *Cancers*, 3(3):3242–3278.

Parsons, D. W., Jones, S., Zhang, X., Lin, J. C., Leary, R. J., Angenendt, P., Mankoo, P., Carter, H., Siu, I. M., Gallia, G. L., Olivi, A., McLendon, R., Rasheed, B. A., Keir, S., Nikolskaya, T., Nikolsky, Y., Busam, D. A., Tekleab, H., Diaz, L. A., Hartigan, J., Smith, D. R., Strausberg, R. L., Marie, S. K., Shinjo, S. M., Yan,

H., Riggins, G. J., Bigner, D. D., Karchin, R., Papadopoulos, N., Parmigiani, G., Vogelstein, B., Velculescu, V. E., and Kinzler, K. W. (2008). An integrated genomic analysis of human glioblastoma multiforme. *Science*, 321(5897):1807–1812.

Peng, J., Wang, P., Zhou, N., and Zhu, J. (2009). Partial correlation estimation by joint sparse regression models. *Journal of the American Statistical Association*, 104(486).

Ruano, Y., Ribalta, T., de Lope, Á. R., Campos-Martín, Y., Fiaño, C., Pérez-Magán, E., Hernández-Moneo, J.-L., Mollejo, M., and Meléndez, B. (2009). Worse outcome in primary glioblastoma multiforme with concurrent epidermal growth factor receptor and p53 alteration. *American journal of clinical pathology*, 131(2):257–263.

Schäfer, J. and Strimmer, K. (2005). An empirical bayes approach to inferring large-scale gene association networks. *Bioinformatics*, 21(6):754–764.

Sheng, Z., Li, L., Zhu, L. J., Smith, T. W., Demers, A., Ross, A. H., Moser, R. P., and Green, M. R. (2010). A genome-wide rna interference screen reveals an essential creb3l2-atf5-mcl1 survival pathway in malignant glioma with therapeutic implications. *Nature medicine*, 16(6):671–677.

Shinoura, N., Sakurai, S., Shibasaki, F., Asai, A., Kirino, T., and Hamada, H. (2002). Co-transduction of apaf-1 and caspase-9 highly enhances p53-mediated apoptosis in gliomas. *British journal of cancer*, 86(4):587–595.

Singh, D., Chan, J. M., Zoppoli, P., Niola, F., Sullivan, R., Castano, A., Liu, E. M., Reichel, J., Porrati, P., Pellegatta, S., et al. (2012). Transforming fusions of fgfr and tacc genes in human glioblastoma. *Science*, 337(6099):1231–1235.

Soengas, M. S., Alarcon, R., Yoshida, H., Hakem, R., Mak, T., Lowe, S., et al. (1999). Apaf-1 and caspase-9 in p53-dependent apoptosis and tumor inhibition. *Science*, 284(5411):156–159.

Tchorz, J., Tome, M., Cloëtta, D., Sivasankaran, B., Grzmil, M., Huber, R., Rutz-Schatzmann, F., Kirchhoff, F., Schaeren-Wiemers, N., Gassmann, M., et al. (2012). Constitutive notch2 signaling in neural stem cells promotes tumorigenic features and astroglial lineage entry. *Cell death & disease*, 3(6):e325.

Verhaak, R. G., Hoadley, K. A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M. D., Miller, C. R., Ding, L., Golub, T., Mesirov, J. P., Alexe, G., Lawrence, M., O'Kelly, M., Tamayo, P., Weir, B. A., Gabriel, S., Winckler, W., Gupta, S., Jakkula, L., Feiler, H. S., Hodgson, J. G., James, C. D., Sarkaria, J. N., Brennan, C., Kahn, A., Spellman, P. T., Wilson, R. K., Speed, T. P., Gray, J. W., Meyerson, M., Getz, G., Perou, C. M., and Hayes, D. N. (2010). Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell*, 17(1):98–110.

Weber, G. L., Parat, M.-O., Binder, Z. A., Gallia, G. L., and Riggins, G. J. (2011). Abrogation of pik3ca or pik3r1 reduces proliferation, migration, and invasion in glioblastoma multiforme cells. *Oncotarget*, 2(11):833.

Yu, H., Liu, B.-H., Ye, Z.-Q., Li, C., Li, Y.-X., and Li, Y.-Y. (2011). Link-based quantitative methods to identify differentially coexpressed genes and gene pairs. *BMC bioinformatics*, 12(1):315.