Appendix

The Appendix contains additional numerical study results.

ANALYSIS OF THE TCGA GBM DATA (THE APOPTOSIS PATHWAY): HISTOGRAMS OF TWO CNAS

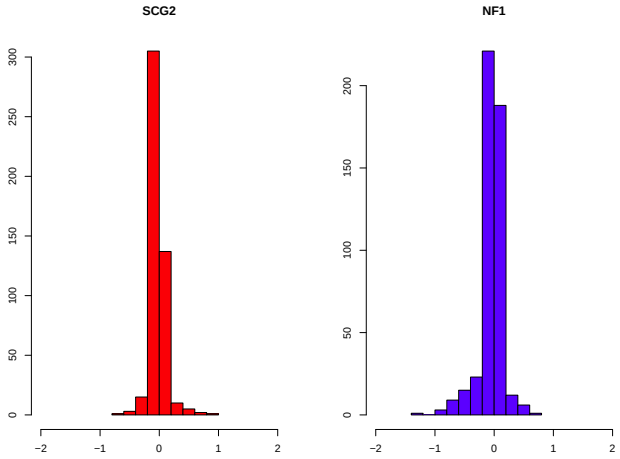

Fig. 3. Analysis of TCGA GBM data: histograms of two CNAs.

ADDITIONAL SIMULATION RESULTS

Table 4. Simulation study of Example 1 with 40 true positives. In each cell, the three rows are model error, number of true positives, and number of false positives. Mean (standard deviation).

Correlation	AR				Block			
(ρ_1, ρ_2)	P_0	P_C	P_G	P_{GC}	P_0	P_C	P_G	P_{GC}
	1.36(0.74)	1.31(0.74)	1.11(0.74)	1.11(0.74)	0.80(0.49)	0.80(0.49)	0.65(0.42)	0.64(0.43)
(0.1, 0.1)	38.1(3.6)	38.2(3.6)	38.9(3.8)	38.9(3.8)	39.2(3.5)	39.2(3.5)	39.6(3.6)	39.6(3.6)
	1.7(2.6)	1.7(2.7)	2.0(3.4)	2.0(3.4)	0.8(1.0)	0.8(1.0)	0.8(1.0)	0.8(1.0)
	1.12(0.69)	1.12(0.69)	0.95(0.54)	0.93(0.51)	0.77(0.52)	0.77(0.52)	0.64(0.40)	0.63(0.41)
(0.1, 0.9)	37.8(4.2)	37.8 (4.2)	38.4(4.2)	38.4(4.2)	38.6(4.0)	38.6(4.0)	38.9(4.1)	38.9(4.1)
	1.5(1.9)	1.5(1.9)	2.0(1.9)	2.0(2.0)	0.8(1.1)	0.8(1.1)	0.7(1.0)	0.8(1.1)
	0.88(0.27)	0.56(0.18)	0.61(0.24)	0.52(0.16)	0.52(0.16)	0.43(0.12)	0.43(0.11)	0.42(0.10)
(0.5, 0.1)	38.9(3.6)	40.0(3.4)	40.0(3.5)	40.2(3.4)	38.9(4.0)	39.3(4.0)	39.3(4.0)	39.4(4.0)
	0.7(1.1)	0.5(0.8)	1.2(1.3)	0.9(0.9)	0.3(0.5)	0.4(0.7)	0.5(0.9)	0.8(1.0)
	0.73(0.39)	0.53(0.24)	0.59(0.32)	0.51(0.24)	0.56(0.35)	0.46(0.22)	0.51(0.30)	0.47(0.25)
(0.5, 0.9)	37.7(4.5)	38.5(4.3)	38.2(4.4)	38.6(4.2)	39.2(4.2)	39.8(3.7)	39.6(4.0)	39.8(3.7)
	0.6(0.8)	0.8(0.9)	0.9(1.3)	1.0(1.2)	0.2(0.5)	0.6(1.0)	0.7(1.4)	0.8(1.2)
	4.26(1.08)	0.68(0.19)	1.74(0.92)	0.72(0.22)	4.87(0.86)	0.65(0.15)	2.60(0.98)	0.68(0.17)
(0.9, 0.1)	13.4(2.2)	35.0(3.5)	26.8(6.5)	34.5(3.2)	10.0(0.5)	34.1(3.9)	21.6(4.8)	33.7(3.8)
	2.1(2.6)	4.2(2.9)	6.9(3.3)	4.0(2.9)	0.2(0.5)	3.5(2.1)	3.4(2.7)	2.7(1.6)
	4.12(0.82)	0.76(0.26)	1.75(0.82)	0.76(0.22)	4.93(0.67)	0.66(0.22)	2.28(1.00)	0.66(0.22)
(0.9, 0.9)	12.6(2.4)	33.7(4.3)	26.1(4.7)	34.2(3.9)	10.0(0.6)	34.8(3.3)	24.1(5.6)	34.5(3.6)
	2.5(2.6)	5.1(3.1)	6.2(2.8)	6.5(5.6)	0.3(0.5)	3.6(1.3)	2.3(2.0)	3.5(1.5)

Table 5. Simulation study of Example 2 with 40 true positives. In each cell, the three rows are model error, number of true positives, and number of false positives. Mean (standard deviation).

Correlation	AR				Block			
(ρ_1, ρ_2)	P_0	P_C	P_G	P_{GC}	P_0	P_C	P_G	P_{GC}
	2.59(1.15)	2.57(1.16)	2.20(1.15)	2.20(1.15)	2.43(1.09)	2.43(1.10)	1.99(0.91)	1.99(0.92)
(0.1, 0.1)	36.4(4.1)	36.4(4.06)	37.5(4.4)	37.5(4.4)	35.3(4.3)	35.3(4.3)	36.5(4.5)	36.5(4.5)
	6.6(6.1)	6.5(6.1)	6.8(6.47)	6.8(6.5)	4.6(4.7)	4.7(4.7)	4.2(4.0)	4.2(4.0)
	2.72(1.47)	2.72(1.47)	2.03(1.22)	2.03(1.22)	2.23(1.44)	2.23(1.45)	1.77(1.25)	1.77(1.25)
(0.1, 0.9)	35.3(4.2)	35.3(4.2)	37.5(3.8)	37.5(3.8)	35.8(5.1)	35.9(5.1)	37.1(4.6)	37.1(4.6)
	5.1(5.8)	5.1(5.8)	6.3(5.4)	6.3(5.2)	3.7(4.7)	3.7(4.6)	4.7(4.0)	4.6(4.0)
	5.61(0.75)	5.53(0.72)	5.51(0.73)	5.49(0.73)	5.64(0.82)	5.48(0.77)	5.45(0.81)	5.38(0.81)
(0.5, 0.1)	13.5(3.0)	14.1(2.8)	14.3(2.7)	14.4(2.8)	17.7(4.1)	19.1(4.2)	19.2(4.6)	19.8(4.7)
	2.2(3.3)	2.2(2.3)	2.4(3.2)	2.4(2.4)	2.0(2.9)	2.0(3.1)	2.2(3.5)	2.2(3.6)
	5.55(0.75)	5.51(0.76)	5.09(0.68)	5.07(0.69)	5.80(1.06)	5.64(1.01)	5.05(1.01)	4.97(0.99)
(0.5, 0.9)	14.2(4.2)	14.6(4.1)	18.0(4.9)	18.2(5.0)	16.9(4.2)	19.0(4.2)	22.2(5.0)	23.1(5.2)
	2.47(2.5)	2.9(3.3)	5.1(6.4)	5.4(6.4)	1.7(2.0)	2.4(3.2)	3.2(3.9)	3.6(4.0)
	1.75(0.25)	1.50(0.22)	1.61(0.23)	1.47(0.25)	1.77(0.24)	1.27(0.35)	1.64(0.24)	1.26(0.33)
(0.9, 0.1)	6.8(2.0)	12.4(4.5)	8.2(3.0)	13.0(5.0)	6.7(1.8)	19.4(6.8)	8.5(3.0)	19.6(6.5)
	1.6(1.6)	4.8(3.2)	1.9(1.7)	4.9(3.2)	0.5(0.8)	2.4(1.9)	0.7(0.8)	2.6(2.1)
	1.69(0.38)	1.43(0.39)	1.43(0.32)	1.35(0.33)	1.81(0.29)	1.29(0.32)	1.50(0.27)	1.21(0.26)
(0.9, 0.9)	7.7(2.6)	14.4(5.8)	12.6(4.1)	15.4(5.5)	6.7(1.3)	20.0(5.6)	12.4(3.7)	20.7(5.8)
	1.7(1.8)	5.7(4.4)	3.6(2.9)	4.7(2.9)	0.2(0.6)	2.6(1.7)	1.4(1.4)	2.5(2.0)

ANALYSIS OF THE TCGA GBM DATA (THE APOPTOSIS PATHWAY) USING P_{UNI}

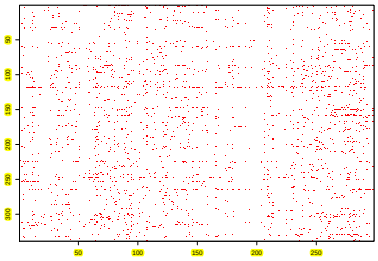

Fig. 4. Positions of the nonzero coefficients under P_{uni}.

ANALYSIS OF THE TCGA GBM DATA (THE APOPTOSIS PATHWAY): REVERSED REGRESSION

In the main text, GEs are regressed on CNAs. The regression model has been motivated by the fact that CNAs regulate the levels of GEs. One reviewer suggested that, to complement the existing analysis, it is of interest to conduct the "reversed" regression of CNAs on GEs. This analysis is conducted in a similar manner as the proposed P_{GC}, only with the roles of GE and CNA reversed. This analysis identifies a total of 5,548 associations, with 696 overlapped with the proposed P_{GC}. More detailed results are available from the authors. The observation is that it and P_{GC} lead to significantly different conclusions. With the regulation relationship, we believe that the proposed analysis with GEs regressed on CNAs is more sensible.

ANALYSIS OF THE TCGA GBM DATA (THE CELLULAR LOCALIZATION PATHWAY)

For the GBM data described in the main text, we also analyze the cellular localization pathway, which contains 315 GEs and 368 CNAs. The cellular localization process takes place at the cellular level. As a result of the cellular localization process, a substance or cellular entity

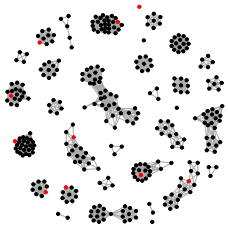

Fig. 5. Networks of CNAs in the apoptosis pathway in GBM. Red dots correspond to CNAs with nonzero coefficients for the expression of gene PTEN identified by P_{uni}.

(such as a protein complex or organelle) is transported to and/or maintained in a specific location within or in the membrane of a cell. This pathway is not cancer related, and this set of analysis can serve as a "negative control" for that in the main text.

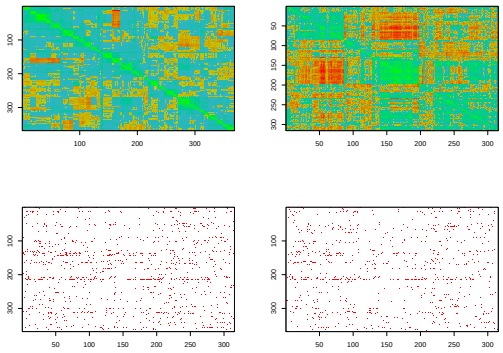

Fig. 6. Analysis of the cellular localization pathway in GBM. Left-upper: heatmap for the correlation matrix of CNAs. Right-upper: heatmap for the correlation matrix of GEs. Left-lower: positions of the nonzero coefficients under P_{GC}. Right-lower: positions of the nonzero coefficients under P_0.

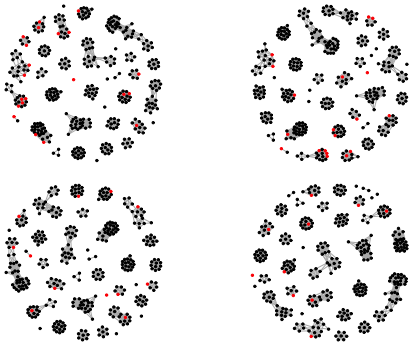

Fig. 7. Networks of CNAs in the cellular localization pathway in GBM. Red dots correspond to CNAs with nonzero coefficients for the expression of gene TRAM1. Left-upper: P_{GC}. Right-upper: P_C. Left-lower: P_G. Right-lower: P_0.

Analysis is conducted in the same manner as in the main text. Figure 6 contains the heatmaps of the correlation coefficient matrices for both CNAs and GEs. The subgraphs for CNAs are further shown in Figure 7. For the 315×368 regression coefficient matrix β, 3930 (marginal),

3423 (P_{uni}), 3898 (P_{GC}), 2903 (P_0), 3868(P_C), and 3572 (P_G) nonzero elements are identified. Different methods identify different sets of associations. For example, there are only 121 overlapped associations between the marginal approach and P_{GC}. For P_{GC} and P_0, we also show in Figure 7 the positions of nonzero regression coefficients (represented by red pixels). We sum over the off-diagonal elements in the correlation matrices of regression coefficients and obtain 149.4 for P_{GC} and 100.8 for P_0, which suggests a higher level of similarity for the proposed method. More detailed results are available from the authors.

As a representative example, we take a closer look at the TRAM1 gene. The identified CNAs and their estimated regression coefficients are shown in Table 6. Note that although the marginal approach also conducts regression, its regression coefficients are not directly comparable to those under the penalization methods. Thus, we use "+" to represent an identified association of the marginal approach. Under all five penalization methods, CNA TRAM1 has positive regression coefficients. Under P_0, P_C, P_G, and P_{GC}, their magnitudes are the largest. It is interesting to note that the marginal approach misses this CNA. In Figure 7, we also show the CNAs identified using different methods with respect to the subgraphs. For the proposed method, seven subgraphs contain at least two identified CNAs. In contrast, no such subgraph is identified by P_0.

ANALYSIS OF THE TCGA LIHC DATA

We analyze the TCGA LIHC (liver hepatocellular carcinoma) data. Data collection and processing are similar to those for the GBM data. Details are omitted here. We analyze the gluconeogenesis pathway. Gluconeogenesis is a metabolic pathway that results in the generation of glucose from non-carbohydrate carbon substrates such as pyruvate, lactate, glycerol, and glucogenic amino acids. In mammals, this process is restricted to liver and kidney. The analyzed data contain 67 GE and 67 CNA measurements on 262 samples. As the previous example, this set of data analysis also serves as a "negative control" for that in the main text.

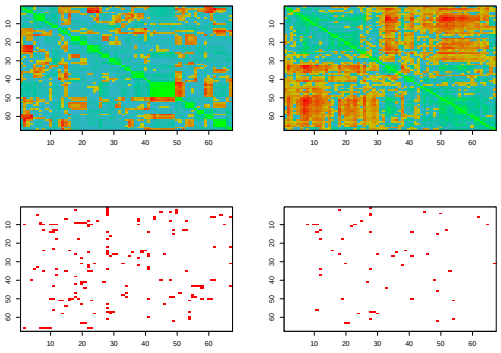

Fig. 8. Analysis of the gluconeogenesis pathway in LIHC. Left-upper: heatmap for the correlation matrix of CNAs. Right-upper: heatmap for the correlation matrix of GEs. Left-lower: positions of the nonzero coefficients under P_{GC}. Right-lower: positions of the nonzero coefficients under P_0.

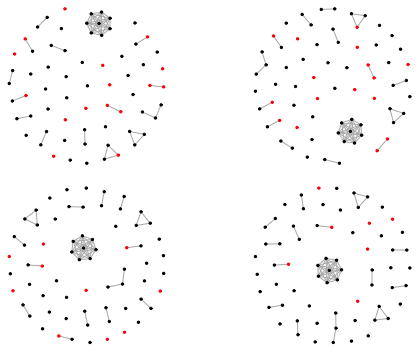

Fig. 9. Networks of CNAs in the gluconeogenesis pathway in LIHC. Red dots correspond to CNAs with nonzero coefficients for the expression of gene LDHAL6A. Left-upper: P_{GC}. Right-upper: P_C. Left-lower: P_G. Right-lower: P_0.

Analysis is conducted in the same manner as with the previous two examples. Figure 8 contains the heatmaps of the correlation coefficient matrices for both CNAs and GEs. The subgraphs for CNAs are further shown in Figure 9. For the 67×67 regression coefficient matrix β, 131 (marginal), 520 (P_{uni}), 171 (P_{GC}), 54 (P_0), 131(P_C), and 89 (P_G) nonzero elements are identified. Different methods identify different sets of associations. For example, there are only 28 overlapped associations between the marginal approach and P_{GC}. For P_{GC} and P_0, we also show in Figure 9 the positions of nonzero regression coefficients (represented by red pixels). We sum over the off-diagonal elements in the correlation matrices of regression coefficients and obtain 26.2 for P_{GC} and 1.5 for P_0, which suggests a higher level of similarity for the proposed method.

As in the previous two analyses, we choose a representative gene, LDHAL6A, and take a closer look. The identified CNAs and their estimated regression coefficients are shown in Table 7. The observations are somewhat different from the previous analyses. Specifically, under P_G and P_{GC}, CNA LDHAL6A has negative regression coefficients. The marginal approach, P_0, and P_C miss this CNA. P_{uni} returns a null model. Such results suggest a weaker signal of CNA LDHAL6A on the corresponding gene expression. In Figure 9, we show the CNAs identified using different methods with respect to the subgraphs. For P_{GC}, two subgraphs contain at least two identified CNAs. In contrast, no such subgraph is identified by P_0.

Table 6. Regression coefficients for the expression of gene TRAM1 in the cellular localization pathway in GBM using different methods. For the marginal approach, a "+" represents an identified association.

	marginal	P_{uni}	P_0	P_C	P_G	P_{GC}
ABCG1				-0.355		-0.348
AGXT		0.723				
AP3B2		0.132		0.326		0.307
APOE				0.176		0.183
ATXN1		-1.564				
BAIAP3			-0.367		-0.346	-0.109
BCL3		0.301				
BIRC5		0.256				
CCL3			-0.734	-0.443	-0.706	-0.453
CCL8				0.424		0.425
CIDEA		-0.049				
COPZ1				0.169		0.170
CRYAA		0.149	0.276	0.609	0.284	0.604
DNAJC1		-0.175	-0.412	-0.163	-0.378	-0.164
DOPEY1		0.311				
EGFR		0.076	0.089	0.053	0.084	0.054
ERP29		-0.120				
FAF1		0.026				
GLMN		-0.208				
GOLGA5	+					
HPS4		-0.101				
KLHL2		0.344	0.465	0.436	0.460	0.437
KPNA3		-0.151				
MON2		0.326	0.309	0.227	0.295	0.226
MYO1E			0.335	0.065	0.318	0.082
NF1				-0.318		-0.313
NIN		-0.024				
NLGN1	+					
NLRC4	+					
NOP58		-1.420				
NPM1		0.316				
NUP133	+					
PDIA2		-0.025		-0.307		-0.217
PDIA4		-0.001				
PEX13	+					
PEX14		-0.054			-0.161	
PPY		-0.911		-0.460		-0.454
RPH3AL	+					
RPGR		-0.550				
SCG5		0.273				
SCRN1		0.113	0.114	0.106	0.134	0.106
SMG5		-0.041				
SSR1		1.302				
STX16	+					
STX18		0.187				
TIMM17A	+					
TNFSF14						-0.017
TOPORS		-0.045		-0.191	-0.160	-0.193
TRAM1		0.766	1.096	1.069	1.119	1.071
TPR	+					
VCP		0.766				
ZWINT		-0.148		-0.404		-0.402

Table 7. Regression coefficients for the expression of gene LDHAL6A in the gluconeogenesis pathway in LIHC using different methods. For the marginal approach, a "+" represents an identified association.

	marginal	P_{uni}	P_0	P_C	P_G	P_{GC}
LDAH	+					
HK3			-1.018	-1.348	-1.357	-1.352
HK2			0.137	1.363	1.556	1.360
GPI			0.643	0.692	0.828	0.693
ALDOA			-0.212	-0.519	-0.723	-0.521
ALDH1B1			0.271	1.232	1.072	1.230
ACSS2			1.751	2.065	2.056	2.069
PGM1			0.808	2.013	1.623	2.007
HK1				0.739	1.677	0.746
ALDOC				-0.793	-0.731	-0.854
PGAM1				0.682		0.677
ENO3				-1.106	-0.941	-1.103
ENO1				-0.449		-0.442
LDHAL6A				-0.759		-0.765
AKR1A1				-0.311		-0.313
G6PC3				-0.131		-0.053
MINPP1				0.581		0.584