Supplementary Material for "Bayesian Hierarchical Varying-sparsity Regression Models with Application to Cancer Proteogenomics"

Yang Ni[1], Francesco C. Stingo[2], Min Jin Ha[3], Rehan Akbani[4], and Veerabhadran Baladandayuthapani[3]

[1]Department of Statistics and Data Sciences, The University of Texas at Austin
[2]Department of Statistics, Computer Science, Applications "G. Parenti", The University of Florence
[3]Department of Biostatistics, The University of Texas MD Anderson Cancer Center
[4]Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center

October 3, 2017

A. Additional modeling details

When considering selection of significant spline coefficients, generically denoted by $\boldsymbol{\alpha}$, conventionally, we can assign a spike-and-slab prior,

$$\boldsymbol{\alpha} \sim \gamma \mathrm{N}(0, \Sigma) + (1-\gamma)\delta_0$$

and update $\boldsymbol{\alpha}$ and γ by blocked Gibbs sampler. Notice that each spline coefficient is a vector. This joint update step usually has low acceptance rate because whenever we propose $\gamma : 0 \to 1$, we also need to propose new values of vector $\boldsymbol{\alpha}$ from an independent proposal (e.g. its prior). It is difficult to propose a "good" candidate $\boldsymbol{\alpha}$ if it is multidimensional and therefore the mixing of the Markov chain is quite unsatisfactory. To improve the mixing, we follow the treatment in Gelman et al. (2008) and Scheipl et al. (2012) to expand the vector parameter $\boldsymbol{\alpha} = \eta\boldsymbol{\xi}$ and assign a spike-and-slab prior on the scalar parameter η. Note that updating η and its spike-and-slab indicator γ is straightforward since γ is now conditional independent of multidimensional vector $\boldsymbol{\alpha}$ given η, which can be seen from the

Markov blanket in Figure 2 of this Supplementary Material. Moreover, updating $\boldsymbol{\xi}$ is not difficult either because (i) it is not updated jointly with indicator γ and therefore can be updated with a random-walk proposal, which has a much higher acceptance rate than an independent proposal and (ii) the dimension of $\boldsymbol{\xi}$ is small to moderate (usually around 10) due to the spectral decomposition of our spline basis construction in Section 3 where we retain only the first several eigenvectors and eigenvalues that explain most of the variability of the splines.

We discuss three important factors that characterize the spline: (1) degree of the bases, (2) number of the bases and (3) position of the knots.
(1) Degree r of the bases. In Section 3, we essentially decompose the nonlinear function $f_{jk}(G_{ijk}) = f^0_{jk}(G_{ijk}) + f^{pen}_{jk}(G_{ijk})$ into a polynomial part (f^0_{jk}) and a nonlinear part (f^{pen}_{jk}). Since the degree of polynomial function $f^0_{jk}(\cdot)$ equals $r-2$ (because the penalty matrix $\boldsymbol{K}$ is rank-2 deficient), we choose the degree of the bases to be 3 so that the polynomial function $f^0_{jk}(G_{ijk}) = G_{ijk}\alpha^0_{jk}$ is linear. This allows us to apply variable selection technique separately to each part and to differentiate between linear trend and nonlinear trend. Setting $r = 2$ or 4, however, would not allow us to differentiate between linear and nonlinear parts anymore (Scheipl et al., 2012).
(2) Number b of spline bases. We choose a large b (e.g. 20) to flexibly capture the nonlinearity of $f_{jk}(\cdot)$. To prevent overfitting, we penalize the spline coefficients by a roughness penalty and only keep a few transformed bases (usually around 10) which explain most of the variability of $f_{jk}(\cdot)$ (Ruppert et al., 2003).
(3) Position of the knots. We place the knots on the quantiles of gene expressions so that more bases are used to capture the local structure where there are more data points (Ruppert et al., 2003). Alternatively, we can also use equally spaced knots. The choice between quantile or equally spaced knots does not appear to be an critical decision here due to the large number of knots.

A key feature of the Bayesian hard-thresholding mechanism $\beta = \theta I(|\theta| > \lambda)$ is the randomness in both θ and λ (i.e. they are both assigned prior distributions). For illustration, we consider two toy examples: (1) $\theta \sim N(0.2, 0.3^2)$ and (2) $\theta \sim N(0.2, 1)$. In both cases, θ has mean 0.2 but has very different variances, and $\lambda \sim \text{Unif}(0, 1)$. We perform Monte Carlo simulation to evaluate $p(\beta = 0) = p(|\theta| < \lambda)$. Specifically, we draw θ_i and λ_i from their respective distributions and approximate $p(\beta = 0) \approx \frac{1}{N}\sum_{i=1}^{N} I(|\theta_i| < \lambda_i)$ for $i = 1, \ldots, N$ with $N = 10,000$ samples. The resulting probability $p(\beta = 0)$ equals 0.70 for case (1) v.s. 0.36 for case (2). Alternatively, if we fix the threshold at $\lambda = 0.5$ as depicted in Figure 1, the probability $p(\beta = 0|\lambda = 0.5)$ equals 0.93 for case (1) v.s. 0.38 for case (2). Therefore, not only the magnitude but the variability of the coefficient β are taken into account in the Bayesian thresholding mechanism.

Allowing for randomness in the threshold λ necessitates the elicitation of its prior distribution. We find, in Section C, that BEHAVIOR is relatively insensitive to the hyperparameter b_λ of λ. Here we provide a general guidance of choosing b_λ in practice. Researchers can first choose a smallest effect size β_{min}, depending on the application, that they are willing to consider to be significant and then center the uniform prior of the threshold at β_{min} (i.e. setting $b_\lambda = 2\beta_{min}$) since the threshold is interpreted as minimum effect size. Generally, b_λ should not be too small; otherwise the full model is strongly favored *a priori*. If the data

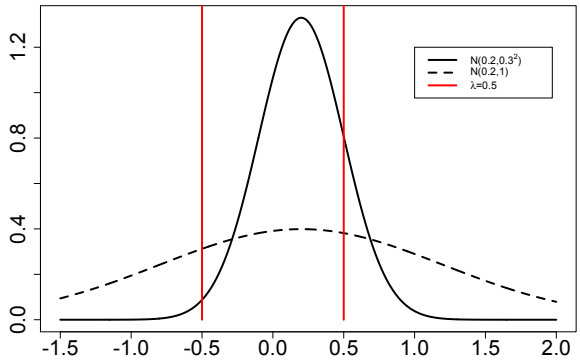

Figure 1: Illustration of the Bayesian thresholding mechanism.

are properly scaled, we recommend to set $b_\lambda \geq 1$, as a reasonable default setting.

A schematic representation of the hierarchical model is shown in Figure 2.

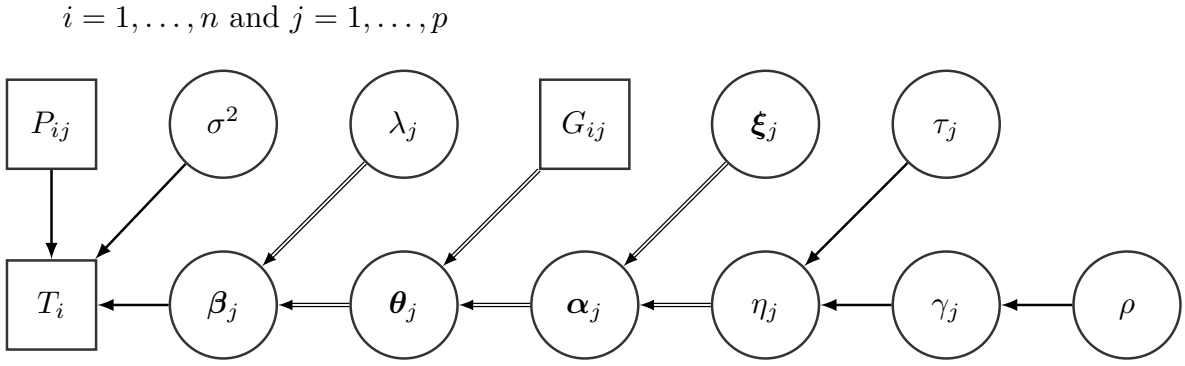

Figure 2: Schema of the Bayesian hierarchical varying-sparsity model. Model parameters are in circles and observed data are in boxes.

B. Posterior inference

In this section, we present the posterior inference procedure under BEHAVIOR and the formula to calculate posterior expected FDR. The MCMC sampler goes as follows.

Algorithm

(1) Update η_{jk} by Metropolis.

(2) Update $\boldsymbol{\xi}_{jk}$ in blocks by Metropolis.

(3) rescale η_{jk} and $\boldsymbol{\xi}_{jk}$: $\boldsymbol{\xi}_{jk} \to \boldsymbol{\xi}_{jk}/\bar{\xi}_{jk}$ and $\eta_{jk} \to \eta_{jk} * \bar{\xi}_{jk}$ with $\bar{\xi}_{jk} = \sum_{i=1}^{d} |\xi_{jk}^{(i)}|/d$.

(4) Update λ by Metropolis.

(5) Update τ_{jk} by Gibbs: $p(\tau_{jk}|\eta_{jk},\gamma_{jk}) = IG(a_\tau + \frac{1}{2}, b_\tau + \frac{\eta_{jk}^2}{2\gamma_{jk}})$.

(6) Update γ_{jk} by Gibbs: $\frac{p(\gamma_{jk}=1|\eta_{jk},\tau_{jk},\rho)}{p(\gamma_{jk}=v_0|\eta_{jk},\tau_{jk},\rho)} = \frac{\sqrt{v_0}\rho}{1-\rho}\exp\left\{\frac{(1-v_0)\eta_{jk}^2}{2v_0\tau_{jk}}\right\}$

(7) Update ρ by Gibbs: $\rho|\boldsymbol{\gamma} \sim Beta(a_\rho + \sum \delta_1(\gamma_{jk}), b_\rho + \sum \delta_{v_0}(\gamma_{jk}))$

(8) Update σ^2 by Metropolis.

Let $\psi_j = I(\beta_j \neq 0)$ denote a binary parameter indicating whether the jth protein is selected in the prognostic model for $j = 1, \ldots, p$. Let $\psi_j^{(i)}$ for $i = 1, \ldots, N$ denote the ith Monte Carlo sample of ψ_j. The expected FDR given a cutoff t is given by

$$FDR(t) = \frac{\sum_{j=1}^p I(\bar{\psi}_j > t)(1-\bar{\psi}_j)}{\sum_{j=1}^p I(\bar{\psi}_j > t)},$$

where $\bar{\psi}_j = \frac{1}{N}\sum_{i=1}^N \psi_j^{(i)}$ is Monte Carlo sample average of ψ_j. Then the desired FDR level can be approximately achieved by choosing an appropriate cutoff t.

C. Sensitivity analysis

We perform a sensitivity analysis on the choice of hyperparameters $(a_\tau, b_\tau) = (5, 100), v_0 = 2.5 \times 10^{-4}, b_\lambda = 1, (a_\sigma, b_\sigma) = (10^{-4}, 10^{-4}), (a_\rho, b_\rho) = (0.5, 0.5)$. We consider the scenario where the true $\lambda = 0.3$ and the results (Tables 1 and 2) show that our approach is relatively robust to the choice of hyperparameters.

Table 1: Sensitivity analysis. The default settings are $(a_\tau, b_\tau) = (5, 100), v_0 = 2.5 \times 10^{-4}, (a_\sigma, b_\sigma) = (10^{-4}, 10^{-4}), (a_\rho, b_\rho) = (0.5, 0.5)$.

	(a_τ, b_τ)	v_0	(a_σ, b_σ)	(a_ρ, b_ρ)
	$(3, 80)$	2.5×10^{-3}	$(10^{-3}, 10^{-3})$	$(0.1, 0.9)$
gTPR	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)
gFDR	0.102 (0.165)	0.268 (0.250)	0.192 (0.245)	0.099 (0.147)
gMCC	0.925 (0.128)	0.778 (0.247)	0.838 (0.242)	0.930 (0.105)
gAUC	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)
pTPR	0.979 (0.025)	0.977 (0.025)	0.976 (0.029)	0.979 (0.025)
pFDR	0.048 (0.047)	0.074 (0.077)	0.056 (0.064)	0.047 (0.047)
pMCC	0.945 (0.043)	0.920 (0.067)	0.935 (0.059)	0.946 (0.044)
pAUC	0.998 (0.003)	0.997 (0.004)	0.998 (0.004)	0.998 (0.003)
MSPE	0.061 (0.027)	0.080 (0.035)	0.065 (0.034)	0.059 (0.027)
	$(8, 120)$	2.5×10^{-5}	$(10^{-5}, 10^{-5})$	$(0.9, 0.1)$
gTPR	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)
gFDR	0.301 (0.290)	0.071 (0.124)	0.185 (0.244)	0.282 (0.303)
gMCC	0.717 (0.334)	0.951 (0.087)	0.844 (0.242)	0.727 (0.348)
gAUC	1.000 (0.000)	0.999 (0.006)	1.000 (0.000)	1.000 (0.000)
pTPR	0.974 (0.028)	0.977 (0.028)	0.976 (0.029)	0.971 (0.035)
pFDR	0.077 (0.088)	0.043 (0.036)	0.056 (0.063)	0.072 (0.079)
pMCC	0.915 (0.076)	0.948 (0.037)	0.936 (0.058)	0.917 (0.071)
pAUC	0.997 (0.004)	0.998 (0.003)	0.998 (0.004)	0.996 (0.005)
MSPE	0.087 (0.043)	0.060 (0.027)	0.065 (0.034)	0.078 (0.040)

Table 2: Sensitivity analysis of b_λ with default $b_\lambda = 1$.

	b_λ					
	0.5	1	1.5	2	5	10
gTPR	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)
gFDR	0.165 (0.236)	0.185 (0.244)	0.193 (0.260)	0.165 (0.225)	0.130 (0.186)	0.169 (0.212)
gMCC	0.862 (0.228)	0.844 (0.242)	0.830 (0.265)	0.864 (0.217)	0.899 (0.177)	0.870 (0.180)
gAUC	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)
pTPR	0.987 (0.017)	0.976 (0.029)	0.976 (0.029)	0.975 (0.031)	0.978 (0.026)	0.978 (0.025)
pFDR	0.066 (0.067)	0.056 (0.063)	0.059 (0.067)	0.057 (0.065)	0.050 (0.060)	0.055 (0.049)
pMCC	0.935 (0.063)	0.936 (0.058)	0.932 (0.062)	0.934 (0.060)	0.942 (0.057)	0.938 (0.042)
pAUC	0.998 (0.004)	0.998 (0.004)	0.997 (0.004)	0.998 (0.004)	0.998 (0.004)	0.998 (0.003)
MSPE	0.065 (0.037)	0.065 (0.034)	0.067 (0.033)	0.064 (0.030)	0.062 (0.029)	0.064 (0.027)

D. Additional details, tables, figures and results for TCGA data analysis

The sample size, the number of events, the median survival time and the average number of selected protein markers per patient per pathway for each cancer is summarized in Table 3. In Figure 3, we show the heatmap of prognostic protein effects $\beta_j(\boldsymbol{G}_{ij})$ for HNSC. In Table 4, we list the proteins and their coding genes (grouped by pathways) that we use in our analysis. In Table 5, we report the c-indices across cancers and pathways.

We also assess the aggregate prognostic effect of each protein. Let

$$\pi_j = \frac{\sum_{i=1}^n p(\boldsymbol{G}_{ij}) p(\beta_j(\boldsymbol{G}_{ij}) \neq 0 | \text{Data})}{\sum_{i=1}^n p(\boldsymbol{G}_{ij})},$$

for $j = 1, \ldots, p$ where $p(\boldsymbol{G}_{ij})$ is the empirical density of gene expressions $\boldsymbol{G}_{ij}$. Intuitively, π_j is the "population-level" marginal posterior probability that protein j is prognostic with patient-level gene expressions $\boldsymbol{G}_{ij}$ integrated out. A large value of π_j suggests that protein j is likely to be prognostic in the population. We reported π_j of each protein for 12 pathways in KIRC in Table 6. There are 18 out of 86 proteins for which the "population-level" marginal posterior probabilities exceed 0.5.

For MCMC convergence diagnostics, we calculate Gelman-Rubin potential scale reduction factor (PSRF, Gelman and Rubin 1992) for continuous parameters and Pearson correlation coefficient of posterior probabilities for binary parameters. The chain is likely to converge to its stationary region when both measures are close to one.

The median PSRFs for $\lambda, \sigma, \alpha, \alpha^0, \alpha^*$ are 1.0010, 1.0002, 1.0002, 1.0003 and 1.0002, respectively. The median correlations for $\gamma_\alpha, \gamma_{\alpha^0}, \gamma_{\alpha^*}, I(\beta = 0)$ are 1.00, 0.94, 0.97 and 1.00, respectively. The trace plots of the log likelihood and the parameters $\lambda, \alpha, \alpha^0, \alpha^*$ for DNA damage response pathway in KIRC are shown in Figures 4-6. The plots of posterior probabilities for binary parameters $\gamma_\alpha, \gamma_{\alpha^0}, \gamma_{\alpha^*}, I(\beta = 0)$ from each pair of chains (totally $\binom{4}{2} = 6$ pairs) are displayed in Figure 7-12.

Table 3: The sample size, the number of events, the median survival time in days and the average number of selected protein markers per patient per pathway for KIRC, OVCA, SKCM and HNSC.

	Sample size	Number of events	Median survival time	Average number of protein markers
KIRC	428	142	2256	2.360
OVCA	232	135	1354	0.126
SKCM	262	105	2993	0.167
HNSC	200	103	804	0.274

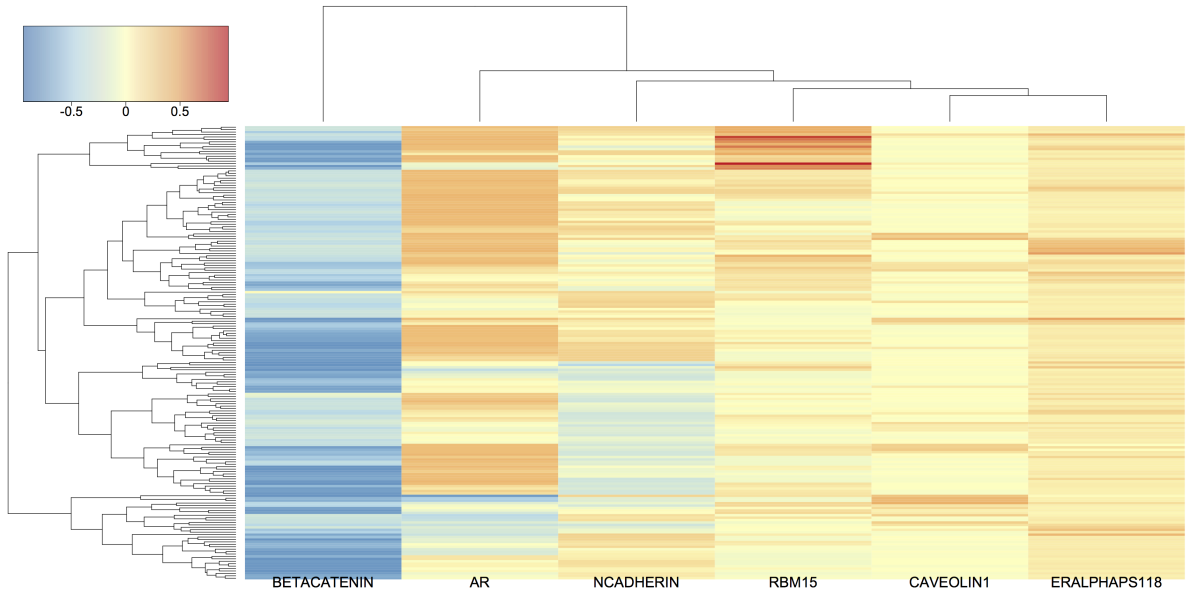

Figure 3: Heatmap of prognostic protein effects $\beta_j(\boldsymbol{G}_{ij})$ for HNSC with PTEN removed. The rows (patients) and columns (proteins) are grouped by hierarchical clustering with complete linkage. Color scale is given in the top left corner.

Table 4: Proteins and corresponding coding genes are listed for 12 pathways.

Pathway	Protein(Gene)
Apoptosis	BAK(BAK1), BAX(BAX), BID(BID), BIM(BCL2L11), BADPS112(BAD), BCL2(BCL2), BCLXL(BCL2L1), CASPASE7CLEAVEDD198(CASP7), CIAP(BIRC2)
Breast reactive	CAVEOLIN1(CAV1), MYH11(MYH11), RAB11(RAB11A,RAB11B), BETACATENIN(CTNNB1), GAPDH(GAPDH), RBM15(RBM15)
Cell cycle	CDK1(CDK1), CYCLINB1(CCNB1), CYCLINE1(CCNE1), PCNA(PCNA), CYCLINE2(CCNE2), P27PT157(CDKN1B), P27PT198(CDKN1B), FOXM1(FOXM1)
Core reactive	CAVEOLIN1(CAV1), BETACATENIN(CTNNB1), RBM15(RBM15), ECADHERIN(CDH1), CLAUDIN7(CLDN7)
DNA damage response	53BP1(TP53BP1), ATM(ATM), BRCA2(BRCA2), CHK1PS345(CHEK1), CHK2PT68(CHEK2), KU80(XRCC5), MRE11(MRE11A), P53(TP53), RAD50(RAD50), RAD51(RAD51), XRCC1(XRCC1)
EMT	FIBRONECTIN(FN1), NCADHERIN(CDH2), COLLAGENVI(COL6A1), CLAUDIN7(CLDN7), ECADHERIN(CDH1), BETACATENIN(CTNNB1), PAI-1(SERPINE1)
Hormone receptor	ERALPHA(ESR1), ERALPHAPS118(ESR1), PR(PGR), AR(AR)
Hormone signaling	INPP4B(INPP4B), GATA3(GATA3), BCL2(BCL2)
PI3K/AKT	AKTPS473(AKT1, AKT2, AKT3), AKTPT308(AKT1, AKT2, AKT3), GSK3ALPHABETAPS21S9(GSK3A, GSK3B), GSK3PS9(GSK3A, GSK3B), P27PT157(CDKN1B), P27PT198(CDKN1B), PRAS40PT246(AKT1S1), TUBERINPT1462(TSC2), INPP4B(INPP4B), PTEN(PTEN)
RAS/MAPK	ARAFPS299(ARAF), CJUNPS73(JUN), CRAFPS338(RAF1), JNKPT183Y185(MAPK8), MAPKPT202Y204(MAPK1,MAPK3), MEK1PS217S221(MAP2K1), P38PT180Y182(MAPK14), P90RSKPT359S363(RPS6KA1), YB1PS102(YBX1)
RTK	EGFRPY1068(EGFR), EGFRPY1173(EGFR), HER2PY1248(ERBB2), HER3PY1298(ERBB3), SHCPY317(SHC1), SRCPY416(SRC), SRCPY527(SRC)
TSC/mTOR	4EBP1PS65(EIF4EBP1), 4EBP1PT37T46(EIF4EBP1), 4EBP1PT70(EIF4EBP1), P70S6KPT389(RPS6KB1), RBPS807S811(RB1) MTORPS2448(MTOR), S6PS235S236(RPS6), S6PS240S244(RPS6)

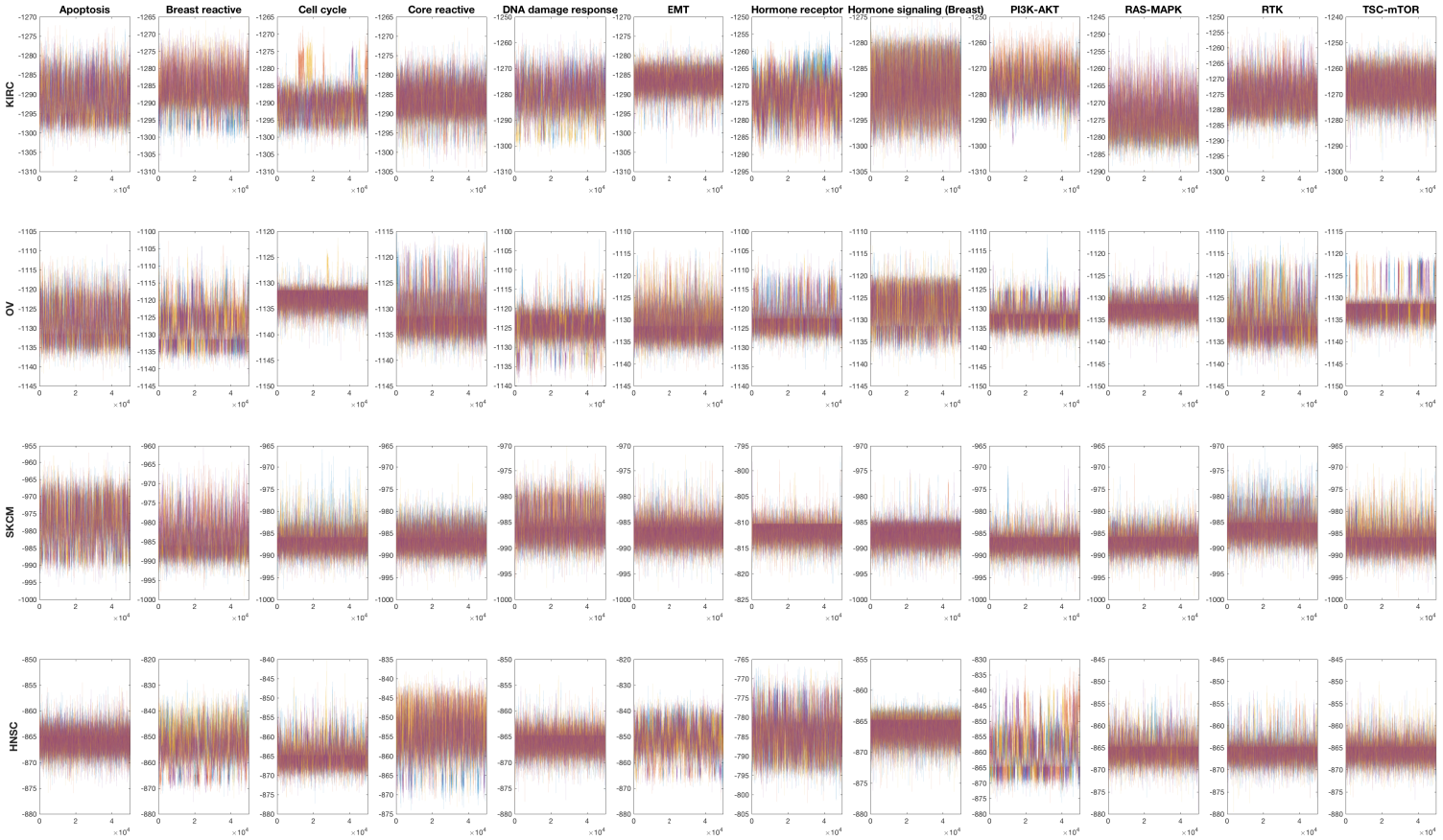

Figure 4: Trace plots of the log likelihood from four parallel chains across 4 cancers and 12 pathways. The x-axes are the iterations and the y-axes are the log likelihood.

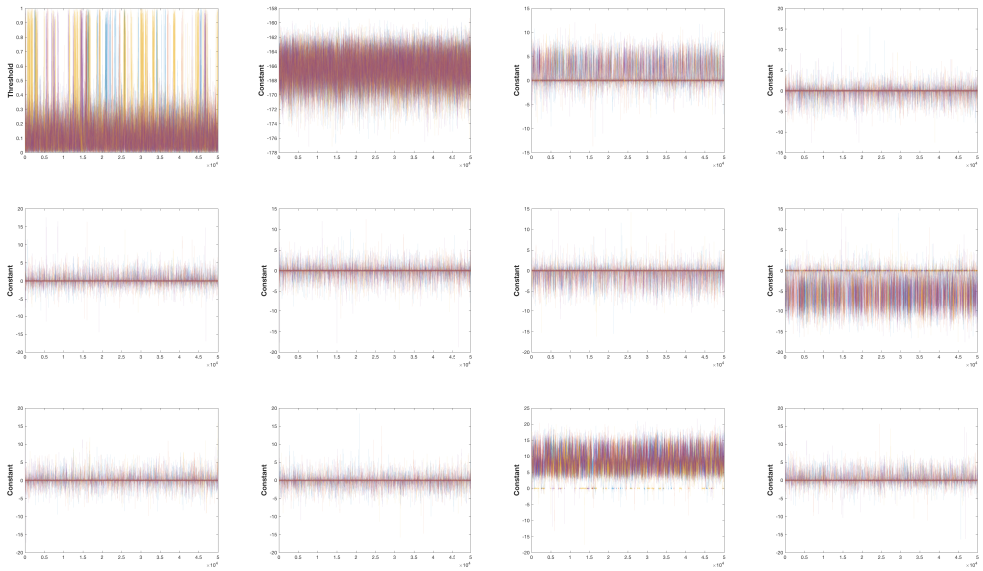

Figure 5: Trace plots of the threshold λ and constant terms α_j from four parallel chains for DNA damage response pathway in KIRC. The x-axes are the iterations. The y-axis of the top left panel is the threshold λ and the y-axes of other panels are the constant terms α_j.

Table 5: C-index of survival time prediction across 4 cancers and 12 pathways. We use "-" to denote the case where BEHAVIOR does not select any protein markers. We follow the acronym of cancer type by TCGA: kidney renal clear cell carcinoma (KIRC), ovarian serous cystadenocarcinoma (OVCA), skin cutaneous melanoma(SKCM) and head and neck squamous cell carcinoma (HNSC). The pathway indices are given by: (1) Apoptosis. (2) Breast reactive. (3) Cell cycle. (4) Core reactive. (5) DNA damage response. (6) EMT. (7) Hormone receptor. (8) Hormone signaling. (9) PI3K/AKT. (10) RAS/MAPK. (11) RTK. (12) TSC/mTOR.

	Pathways											
	1	2	3	4	5	6	7	8	9	10	11	12
KIRC	0.5770	0.6515	0.5698	0.6457	0.6877	0.6084	0.6875	0.5956	0.7049	0.6389	0.6862	0.7224
OVCA	0.5591	0.5791	-	-	0.5165	-	-	0.6167	-	-	-	-
SKCM	0.6080	0.6119	-	-	-	-	-	-	-	-	-	-
HNSC	-	0.6253	-	0.6619	-	0.6576	0.6437	-	0.6305	-	-	-

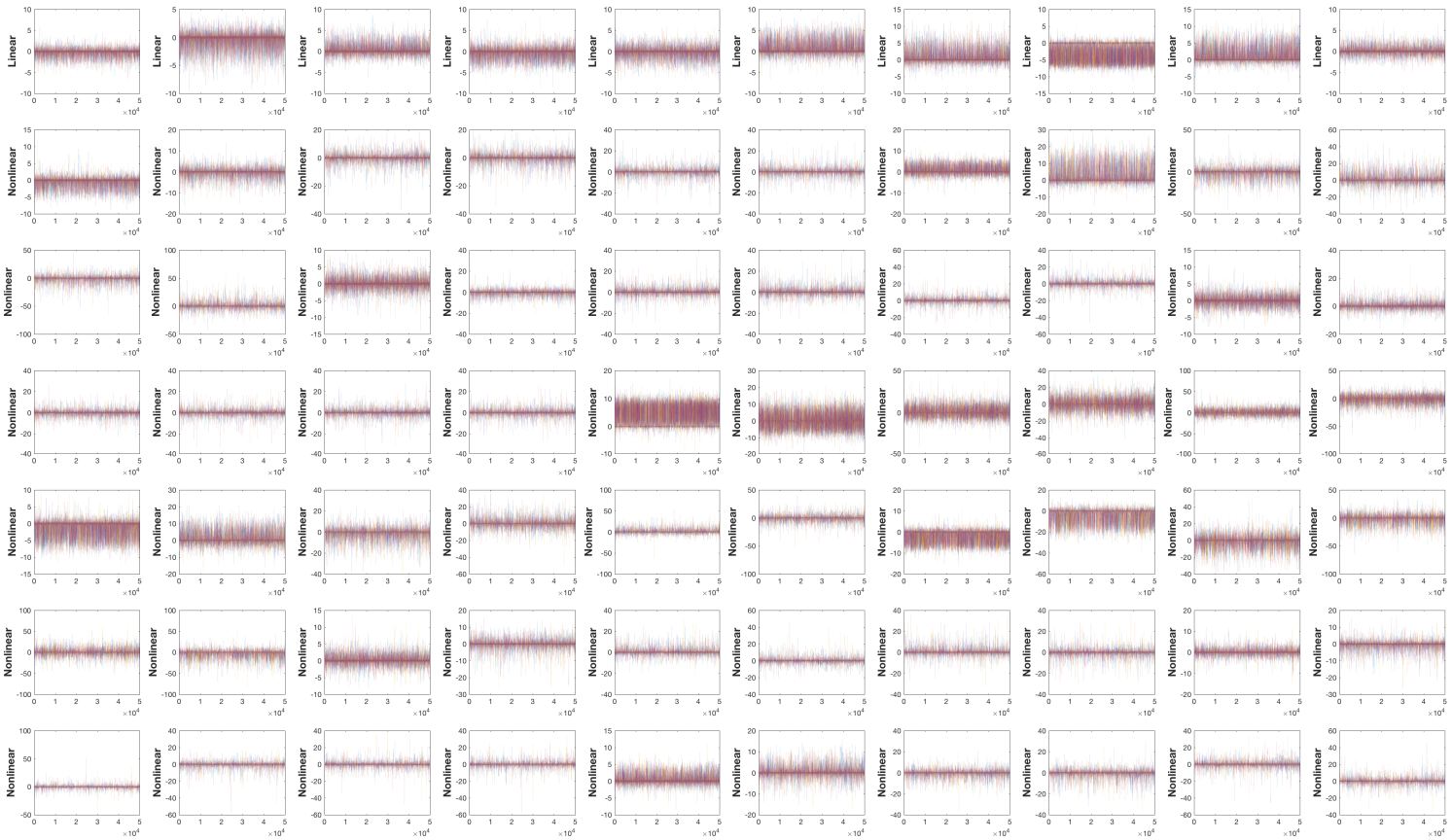

Figure 6: Trace plots of the linear terms α_{jk}^0 and nonlinear terms $\boldsymbol{\alpha}_{jk}^*$ from four parallel chains for DNA damage response pathway in KIRC. The x-axes are the iterations. The y-axes of the panels from the first row are the linear terms α_{jk}^0 and the y-axes of other panels are the nonlinear terms $\boldsymbol{\alpha}_{jk}^*$.

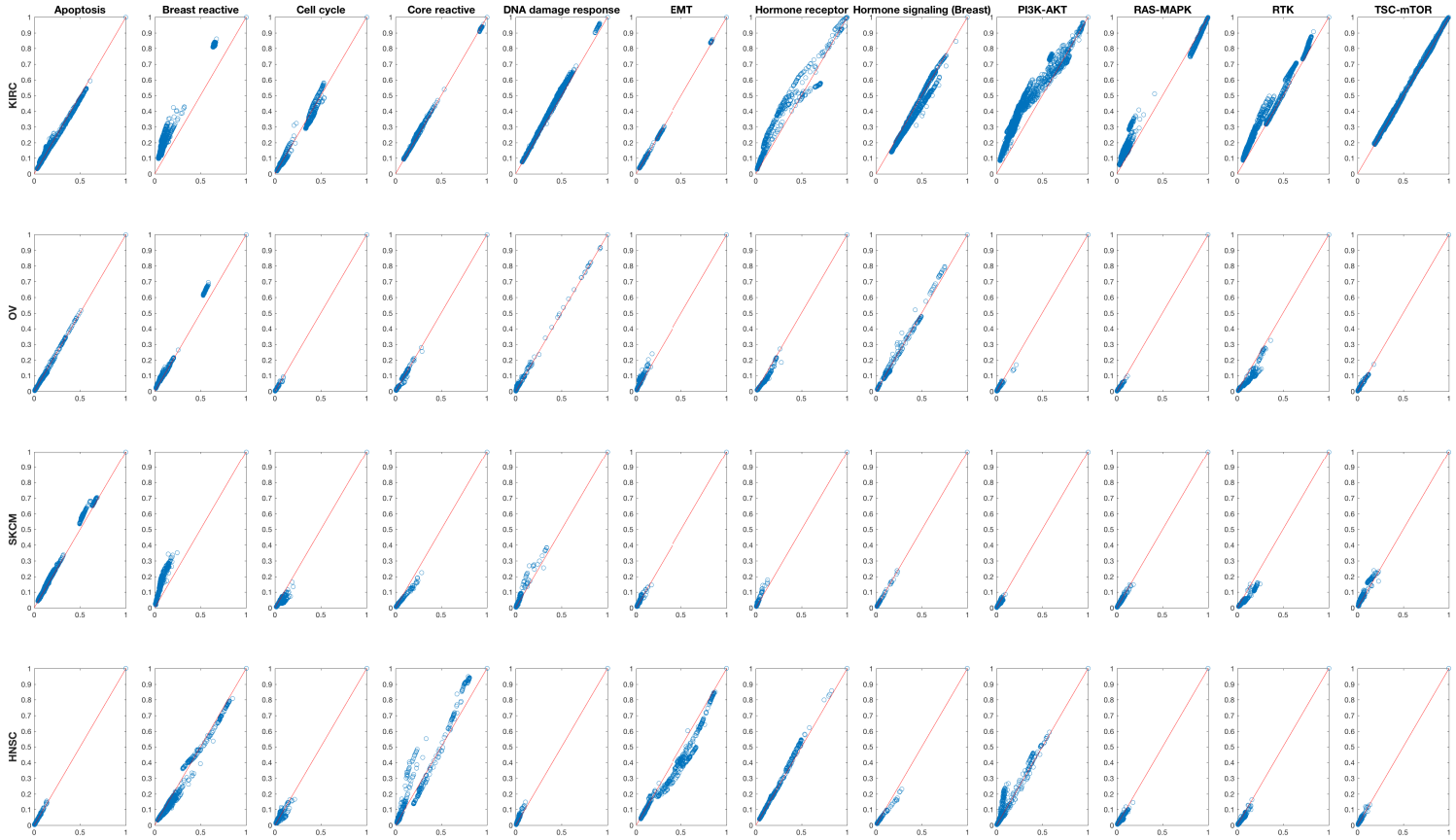

Figure 7: Plots of posterior probabilities for binary parameters $\gamma_\alpha, \gamma_{\alpha^0}, \gamma_{\alpha^*}, I(\beta = 0)$ from the first and second chains across 4 cancers and 12 pathways. The x- and y-axes represent two chains.

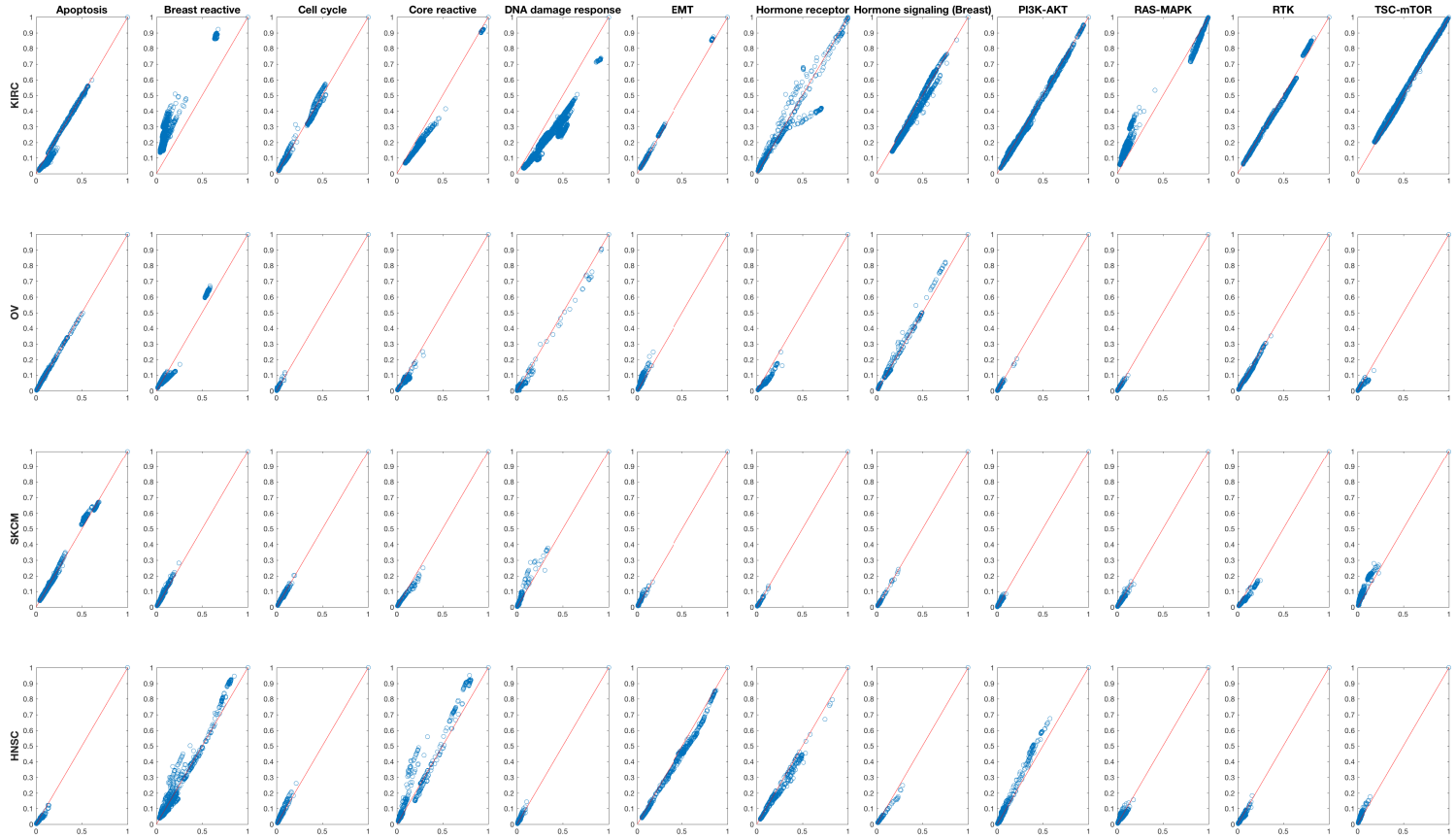

Figure 8: Plots of posterior probabilities for binary parameters $\gamma_\alpha, \gamma_{\alpha^0}, \gamma_{\alpha^*}, I(\beta = 0)$ from the first and third chains across 4 cancers and 12 pathways. The x- and y-axes represent two chains.

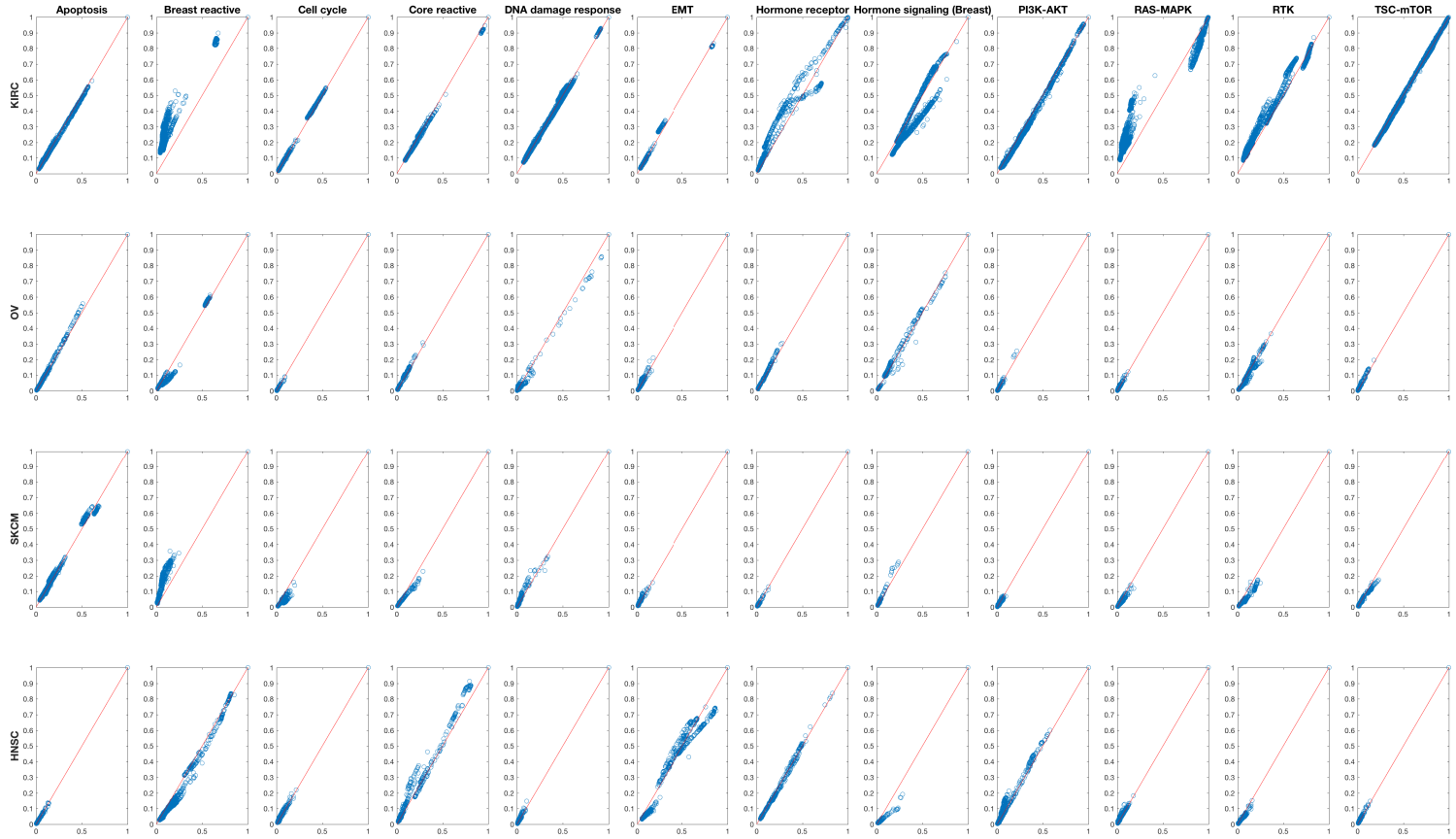

Figure 9: Plots of posterior probabilities for binary parameters $\gamma_\alpha, \gamma_{\alpha^0}, \gamma_{\alpha^*}, I(\beta = 0)$ from the first and fourth chains across 4 cancers and 12 pathways. The x- and y-axes represent two chains.

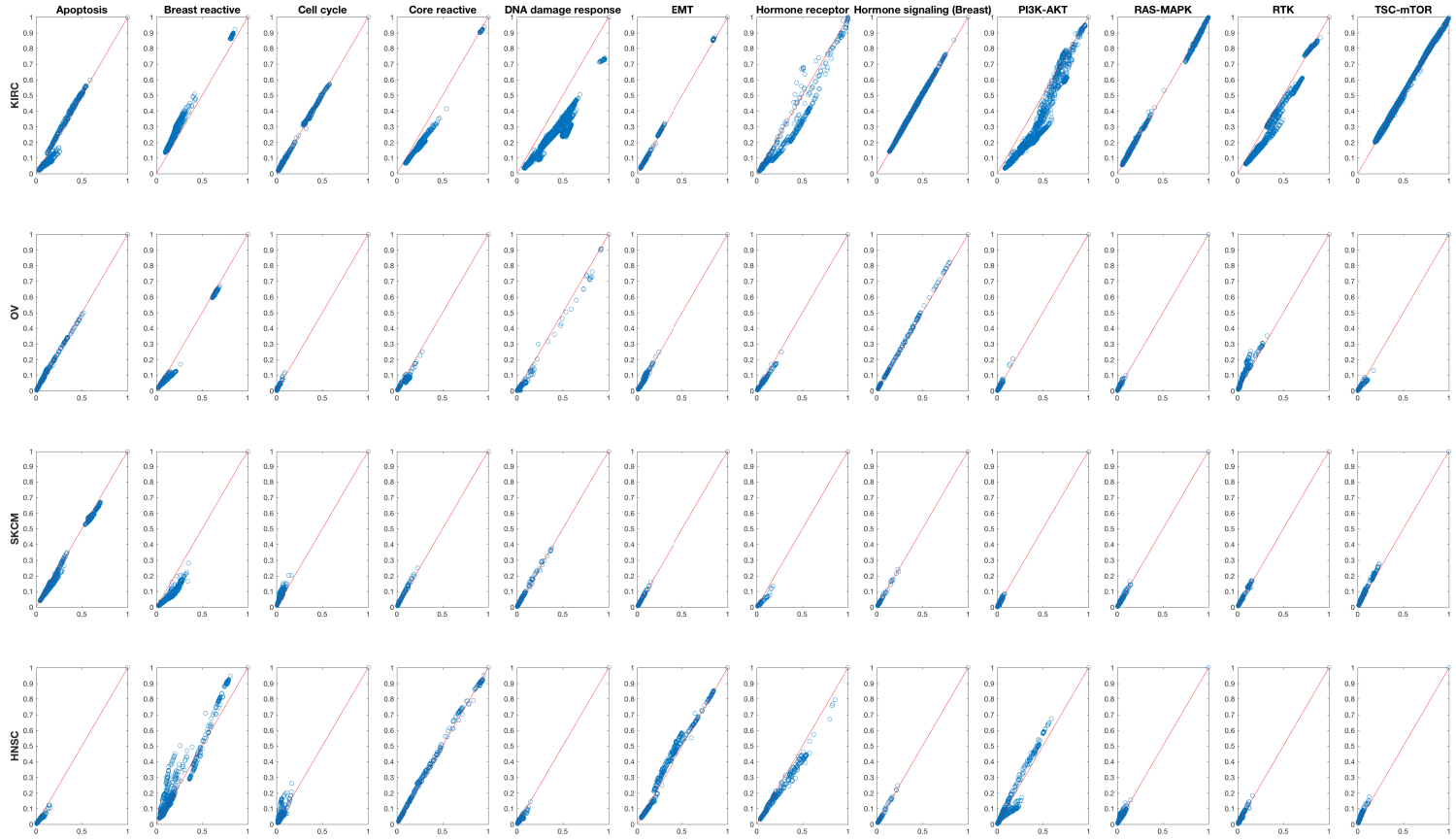

Figure 10: Plots of posterior probabilities for binary parameters $\gamma_\alpha, \gamma_{\alpha^0}, \gamma_{\alpha^*}, I(\beta = 0)$ from the second and third chains across 4 cancers and 12 pathways. The x- and y-axes represent two chains.

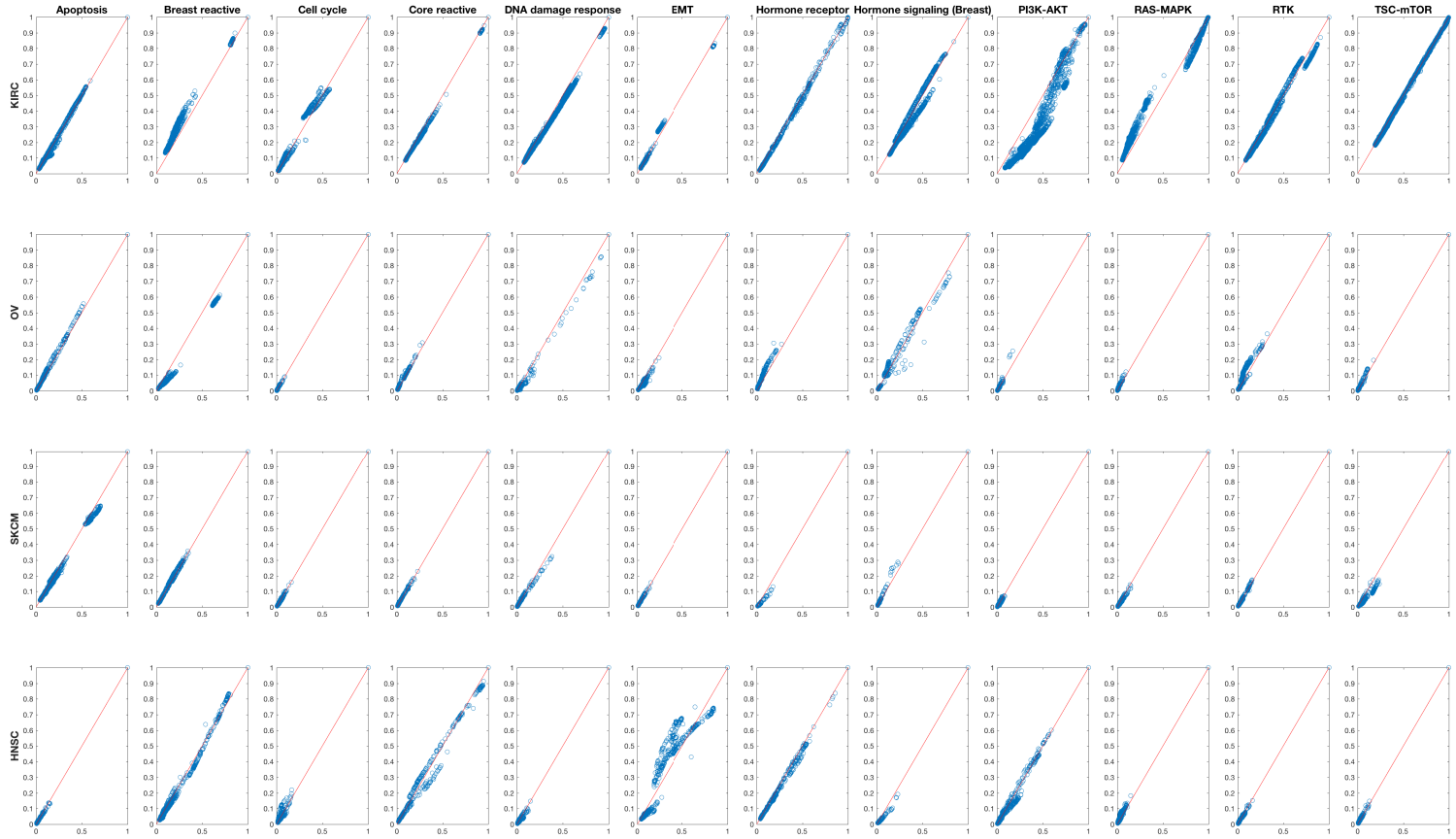

Figure 11: Plots of posterior probabilities for binary parameters $\gamma_\alpha, \gamma_{\alpha^0}, \gamma_{\alpha^*}, I(\beta = 0)$ from the second and fourth chains across 4 cancers and 12 pathways. The x- and y-axes represent two chains.

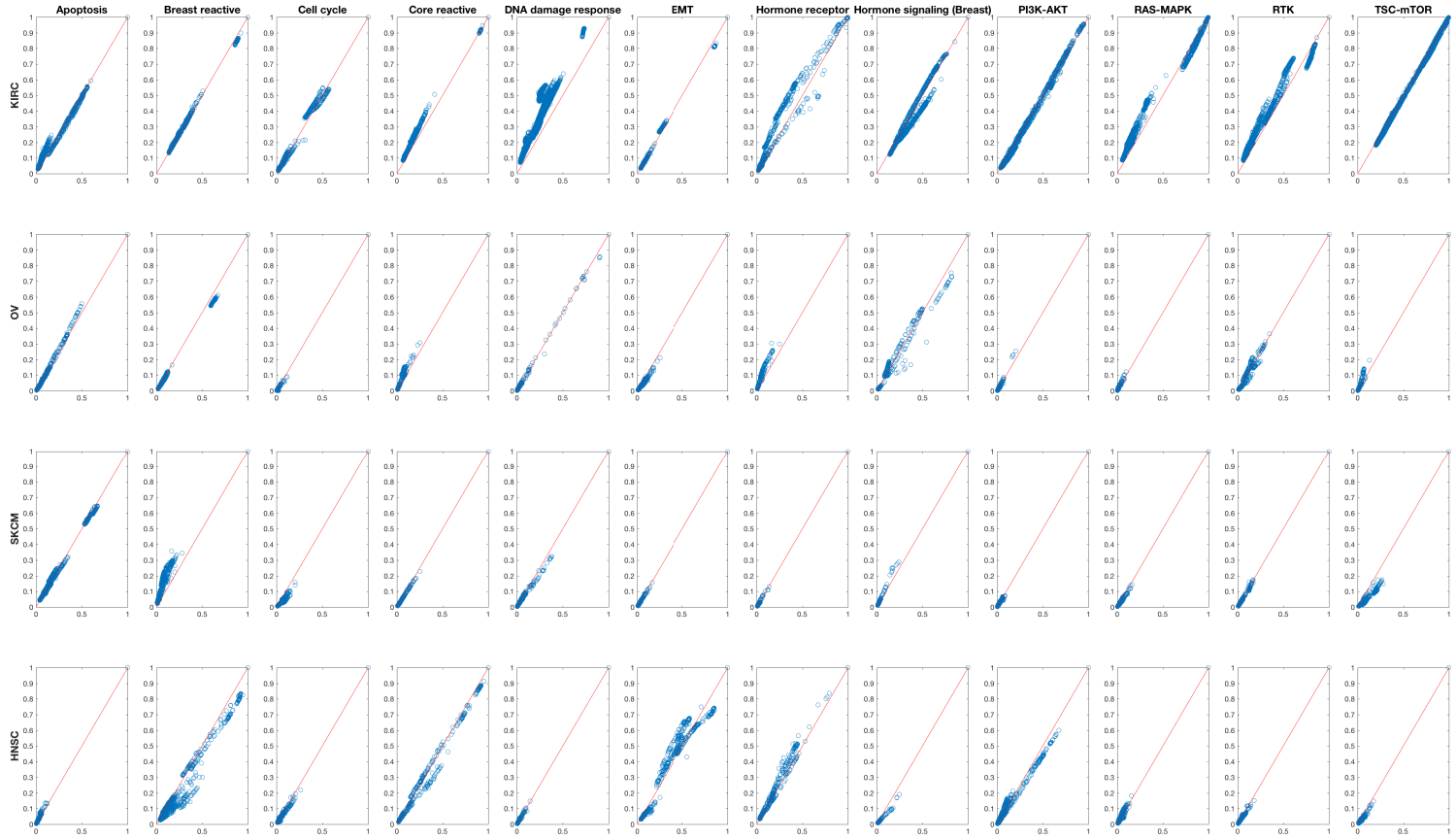

Figure 12: Plots of posterior probabilities for binary parameters $\gamma_\alpha, \gamma_{\alpha^0}, \gamma_{\alpha^*}, I(\beta = 0)$ from the third and fourth chains across 4 cancers and 12 pathways. The x- and y-axes represent two chains.

Table 6: Proteins and their aggregate effects (π_j) are listed for 12 pathways in KIRC. Proteins are bold-faced if $\pi_j > 0.5$.

Pathway	Protein (π_j)
Apoptosis	BAK(0.07), BAX(0.11), BID(0.06), BIM(0.05), BADPS112(0.07), BCL2(0.10), BCLXL(0.41), CASPASE7CLEAVEDD198(0.11), CIAP(0.08)
Breast reactive	CAVEOLIN1(0.36), MYH11(0.25), RAB11(0.31), **BETACATENIN(0.84)**, GAPDH(0.22), RBM15(0.28)
Cell cycle	CDK1(0.08), CYCLINB1(0.47), CYCLINE1(0.09), PCNA(0.09), CYCLINE2(0.21), P27PT157(0.08), P27PT198(0.11), FOXM1(0.08)
Core reactive	CAVEOLIN1(0.29), **BETACATENIN(0.90)**, RBM15(0.29), ECADHERIN(0.17), CLAUDIN7(0.19)
DNA damage response	53BP1(0.43), ATM(0.25), CHK1PS345(0.18), CHK2PT68(0.18), **KU80(0.56)**, **MRE11(0.51)**, P53(0.36), RAD50(0.40), **RAD51(0.95)**, XRCC1(0.18)
EMT	FIBRONECTIN(0.12), NCADHERIN(0.10), COLLAGENVI(0.13), CLAUDIN7(0.09), ECADHERIN(0.08), BETACATENIN(0.32), **PAI-1(0.84)**
Hormone receptor	ERALPHA(0.09), ERALPHAPS118(0.14), PR(0.09), **AR(0.62)**
Hormone signaling	INPP4B(0.44), GATA3(0.25), **BCL2(0.52)**
PI3K/AKT	**AKTPS473(0.57)**, **AKTPT308(0.67)**, GSK3ALPHABETAPS21S9(0.38), GSK3PS9(0.48), **P27PT157(0.52)**, P27PT198(0.26), PRAS40PT246(0.30), **TUBERINPT1462(0.79)**, INPP4B(0.43), PTEN(0.45)
RAS/MAPK	ARAFPS299(0.05), CJUNPS73(0.04), CRAFPS338(0.06), JNKPT183Y185(0.05), **MAPKPT202Y204(0.93)**, MEK1PS217S221(0.11), P38PT180Y182(0.05), P90RSKPT359S363(0.10), YB1PS102(0.14)
RTK	EGFRPY1068(0.22), EGFRPY1173(0.30), HER2PY1248(0.28), HER3PY1298(0.28), **SHCPY317(0.52)**, SRCPY416(0.20), **SRCPY527(0.85)**
TSC/mTOR	4EBP1PS65(0.43), **4EBP1PT37T46(0.73)**, 4EBP1PT70(0.27), **P70S6KPT389(0.83)**, **RBPS807S811(0.60)** MTORPS2448(0.27), S6PS235S236(0.40), S6PS240S244(0.32)

References

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, pages 457–472.

Gelman, A., van Dyk, D. A., Huang, Z., and Boscardin, J. W. (2008). Using redundant parameterizations to fit hierarchical models. *J Comp Graph Stat*, 17(1):95–122.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric regression*. Number 12. Cambridge university press.

Scheipl, F., Fahrmeir, L., and Kneib, T. (2012). Spike-and-slab priors for function selection in structured additive regression models. *Journal of the American Statistical Association*, 107(500):1518–1532.