Supporting Information for Bayesian Data Integration and Variable Selection for Pan-Cancer Survival Prediction using Protein Expression Data

Arnab Kumar Maity [*1], Anirban Bhattacharya [†2], Bani K. Mallick [‡2] and Veerabhadran Baladandayuthapani [§3]

[1]Pfizer, Inc.
[2]Texas A&M University
[3]University of Michigan

Web Appendix A Conditional Distributions and Posterior Computation

In our AFT model for group correlation structure, most of the conditional distributions are available explicitly, hence we can employ Gibbs sampling (Gelfand et al., 1990) technique to explore the posterior distribution. In particular, the complete conditional distributions of β, σ^2, and b_P are given by:

$$\begin{aligned}
\beta|w,\lambda,\tau,\sigma^2 \sim & N(A^{-1}(X^Tw + D^{-1}b_P), \sigma^2A^{-1}) \\
\sigma^2|w,\beta,\lambda,\tau \sim & \text{ Inverse Gamma}\Big(\text{shape} = \frac{n+pr}{2}, \\
& \qquad \text{scale} = \frac{1}{2}\big((w-X\beta)'(w-X\beta) + (\beta-b_P)'D^{-1}(\beta-b_P)\big)\Big) \\
b_{Pj}|\beta_{jk} \sim & N\Big(\frac{1}{\tau_{b_P}\sigma^2\tau^2}\sum_{k=1}^{r}\frac{\beta_{jk}}{\lambda_{jk}^2}, \tau_{b_P}^{-1}\Big), \quad \tau_{b_P} = \frac{1}{\sigma^2\tau^2}\sum_{k=1}^{r}\frac{1}{\lambda_{jk}^2} + \frac{1}{\sigma_P^2}
\end{aligned}$$

where, $A = (X^TX + D^{-1})$, $D = \tau^2\text{diag}\left(\lambda_{11}^2, \ldots, \lambda_{pr}^2\right)$, and $b_P = (b_{P1}, \ldots, b_{Pp})'$.

Due to the nature of the prior on λ and τ, straightforward Gibbs sampling approach may not be possible. An alternative approach, which is based on the idea of slice sampling (Neal,

[*]arnab.maity@pfizer.com
[†]anirbanb@stat.tamu.edu
[‡]bmallick@stat.tamu.edu
[§]veerab@umich.edu

2003), has been discussed in the online supplement of Polson et al. (2014). It follows that,

$$\pi(\lambda_{jk}|\beta_{jk},\tau,\sigma^2) \propto \frac{1}{\lambda_{jk}} \exp\left(-\frac{1}{2}\frac{\beta_{jk}^2}{\lambda_{jk}^2\tau^2\sigma^2}\right)\frac{1}{1+\lambda_{jk}^2} I(\lambda_{jk}>0).$$

Defining $\phi_{jk} = \frac{1}{\lambda_{jk}^2}$ and introducing a latent parameter u_{jk}, the conditional posterior distribution looks like,

$$\pi(u_{jk},\phi_{jk}|\beta_{jk},\tau,\sigma^2) \propto \exp\left(-\frac{1}{2}\frac{\phi_{jk}\beta_{jk}^2}{\tau^2\sigma^2}\right) I(0<u_{jk}<\frac{1}{1+\phi_{jk}}) I(\phi_{jk}>0).$$

Then the following scheme will be used to sample the posterior distribution of λ:

1. Sample $u_{jk}|\phi_{jk} \sim U(0, \frac{1}{1+\phi_{jk}})$.

2. Sample $\phi_{jk}|u_{jk},\beta_{jk},\tau,\sigma^2 \sim$ truncated Exponential$(\frac{\beta_{jk}^2}{2\lambda_j^2\tau^2\sigma^2}) I(0, \frac{1}{u_{jk}}-1)$.

3. Plug back in $\lambda_{jk} = \frac{1}{\sqrt{\phi_{jk}}}$.

Updating τ can be carried out in the similar fashion. We introduce a latent variable v and let $\xi = \frac{1}{\tau^2}$ to yield desired posterior samples:

1. Sample $v|\xi \sim U(0, \frac{1}{1+\xi})$.

2. Sample $\xi|v,\beta_{jk},\lambda_{jk},\sigma^2 \sim$ truncated Gamma$(\frac{pr+1}{2}, \frac{1}{2\sigma^2}\sum_{k=1}^{r}\sum_{j=1}^{p}\frac{\beta_{jk}^2}{\lambda_{jk}^2}) I(0, \frac{1}{v}-1)$.

3. Plug back in $\tau = \frac{1}{\sqrt{\xi}}$.

Finally, we update the censored responses from $w_{ik} \sim N\left(\sum_{j=1}^{p} x_{ijk}\beta_{jk}, \sigma^2\right)$ lower truncated at $\log t_{ik}^*$.

The Markov Chain Monte Carlo (MCMC) chain for simultaneous correlation among groups and proteins (Section 2.3 in the main article) can be constructed extending the strategies detailed above. We implement both types of correlation models along with original horseshoe for a single log normal AFT model in R package **hsaft** and make them available at https://github.com/arnabkrmaity/hsaft/tree/master/hsaft.

Web Appendix B Prediction of Survival Curve

When the interest is to predict the survival time of a new subject having covariate vector x_{new}, the estimation of log survival time proceeds as,

$$\log T_{\text{new}} = \int_w \int_\beta \int_\lambda \int_\tau x_{\text{new}}^T \beta p(w,\beta,\lambda,\tau,\sigma^2|t^*,\delta) dw d\beta d\lambda d\tau d\sigma^2$$

with the corresponding MCMC estimate,

$$\widehat{\log T_{\text{new}}} = \frac{1}{M}\sum_{m=1}^{M} x_{\text{new}}^T \beta^{(m)} | \lambda^{(m)}, \tau^{(m)}, (\sigma^2)^{(m)}$$

where M is the MCMC sample size.

In a very similar manner one obtains the estimated survival probability at time t_0,

$$\begin{aligned}
\widehat{S}(t_0) &= \widehat{\Pr}(T > t_0) \\
&= \frac{1}{M}\sum_{m=1}^{M} \Pr(T > t_0 | x_{\text{new}}, \beta_s^{(m)} \lambda^{(m)}, \tau^{(m)}, (\sigma^2)^{(m)}) \\
&= \frac{1}{M}\sum_{m=1}^{M} \left[1 - \Phi\left((\log t_0 - x_{\text{new}}^T \beta^{(m)}) / (\sigma^2)^{(m)} \right) | \lambda^{(m)}, \tau^{(m)} \right].
\end{aligned}$$

Web Appendix C Impact of the Variance of the mean hyperparameter

To demonstrate that the kidney tumor groups for each protein are correlated, we compute the posterior correlation of β given in (2) using MCMC samples from the first chain. We consider 4 proteins– X1433EPSILON, X4EBP1, FOXO3A_pS318S321, DIRAS3, among which the last two proteins were declared significant by our analysis. Web Table S1 exhibits the computed posterior correlations and one can note that the estimates are indeed high which supports the argument in favor of correlated (or integrated) data analysis. Additionally, we provide the posterior summaries such as trace plots of correlation between tumor groups for the proteins X1433EPSILON, X4EBP1, FOXO3A_pS318S321, and DIRAS3 which indicates of borrowing high correlation in the TCPA data. The plots are provided in Web Figure S2.

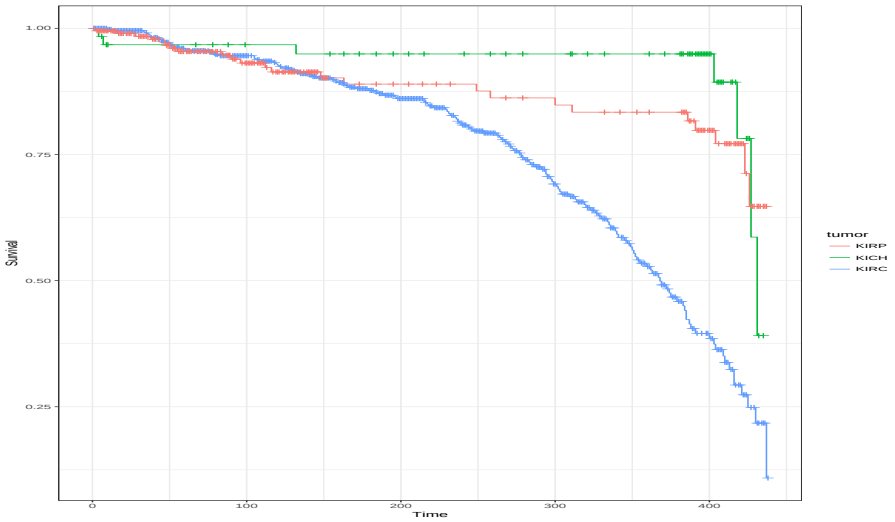

Figure S1: Observed Kaplan-Meier plots for three tumor groups – KICH, KIRC, and KIRP.

Table S1: Posterior correlation of tumor groups for proteins.

Protein	KICH and KIRC	KICH and KIRP	KIRC and KIRP
X1433EPSILON	0.832	0.860	0.842
X4EBP1	0.886	0.892	0.885
FOXO3A_pS318S321	0.727	0.729	0.739
DIRAS3	0.701	0.708	0.725

In the following, we examine the variable selection behavior under the influence of the choice of the variance σ_P^2 in (2) in the main article. To do this, we consider group-corr method in the setting of Example 1 discussed before. We consider several values of the hyperparameter σ_P^2 and compute the area under the ROC curve (AUC). Table S2 reports the results. From these results one can conclude that the variable selection performance is insensitive to the choice of the hyperparameter σ_P^2.

Table S2: AUC under different choices of σ_P^2 in (2) in the main article.

σ_P^2	AUC
0.25	0.638
0.50	0.630
0.75	0.642
1	0.634
5	0.649
10	0.648
15	0.648
20	0.637
25	0.653
30	0.646
40	0.634
50	0.631

Web Appendix D Consistency

In this section, we investigate the frequentist asymptotic behavior of parameter estimates in an AFT log-normal model with a horseshoe prior on the regression coefficients. Bradic et al. (2011) have shown consistency of nonconcave penalized methods for non-polynomial (NP) dimensional data with censoring in the framework of frequentist Cox proportional hazards model. For high-dimensional linear regression models, posterior consistency for the horseshoe prior and its variants have been shown by Armagan et al. (2013). In the context of nonparametric AFT models, Wu and Ghosal (2008) established posterior consistency of the regression function. A combination of these two results provides a consistency result for the regression parameters in the present situation, which is summarized in the following corollary. To be specific, we consider the AFT model prior formulation in (1) for a single tumor group

with $b_{P_j} = 0$, and prove that the posterior distribution concentrates in neighborhoods of the true parameter under certain conditions. The result for any b_{P_j} is beyond the scope of this paper and is due for future research.

Let $\Pi(\beta)$ denote the prior on β and q_n be the number of nonzero elements in β^0 where β^0 is the true value of β. We make use of the following assumptions,

1. $p = o(n)$
2. Let $\Lambda_{n\min}$ and $\Lambda_{n\max}$ be the smallest and the largest singular values of X, respectively, where X denotes the design matrix. Then $0 < \Lambda_{\min} < \liminf_{n\to\infty} \Lambda_{n\min}/\sqrt{n} \leq \limsup_{n\to\infty} \Lambda_{n\max}/\sqrt{n} < \Lambda_{\max} < \infty$
3. $\sup_{j=1,\ldots,rp_n} |\beta_j^0| < \infty$
4. $q_n = o(n/\log n)$

Corollary 1. Under conditions $(1), \ldots, (4)$ and for the horseshoe prior $\Pi(\beta)$, the posterior of β is strongly consistent, that is, for any $\epsilon > 0$, $\Pi(\beta : ||\beta - \beta^0|| > \epsilon|t) \overset{\Pr_{\beta^0}\, a.s.}{\to} 0$ as $n \to \infty$, if

$$\Pi\left(\beta : ||\beta - \beta^0|| < \tfrac{\Delta}{n^{\rho/2}}\right) > \exp(-dn)$$

for all $0 < \Delta < \epsilon^2\Lambda_{\min}^2/(48\Lambda_{\max}^2)$ and $0 < d < \epsilon^2\Lambda_{\min}^2/(32\sigma^2) - 3\Delta\Lambda_{\max}^2/(2\sigma^2)$ and some $\rho > 0$.

Proof. Under the log normal AFT model, for a given ϵ and for a continuous prior $\Pi(\beta)$ on β, Theorem 7.1 of Wu and Ghosal (2008) implies that, as $n \to \infty$,

$$\Pi\left(\beta : ||\beta - \beta^0|| > \epsilon|t\right) \overset{\Pr_{\beta^0}\, a.s.}{\to} 0.$$

In addition, when $\Pi(\beta)$ is the horseshoe prior given in (1) with $b_{P_j} = 0$, then the prior mass condition is verified by Theorem 1 of Armagan et al. (2013), which ensures that for any $\epsilon > 0$, $\Pi(\beta : ||\beta - \beta^0|| > \epsilon|t) \overset{\Pr_{\beta^0}\, a.s.}{\to} 0$ as $n \to \infty$, if

$$\Pi\left(\beta : ||\beta - \beta^0|| < \tfrac{\Delta}{n^{\rho/2}}\right) > \exp(-dn)$$

for all $0 < \Delta < \epsilon^2\Lambda_{\min}^2/(48\Lambda_{\max}^2)$ and $0 < d < \epsilon^2\Lambda_{\min}^2/(32\sigma^2) - 3\Delta\Lambda_{\max}^2/(2\sigma^2)$ and some $\rho > 0$. This completes the proof. □

Web Appendix E Kidney Tumors

We fit the Cox proportional hazard model and the log normal AFT model with randomly selecting 10 covariates on the full dataset. We repeat this procedure 10 times and provide the average AIC and BIC values in Table S3. We notice that goodness of fit criteria such

Table S3: AIC and BIC for Cox proportional hazard model and log normal AFT model for the kidney cancers proteomics data

Model	AIC	BIC
Cox	1895.847	1946.564
AFT	1658.371	1709.089

as AIC and BIC are smaller for the log normal AFT model than those of of the Cox model. This implies that the AFT model provides a better fit at least on the subsets of the data.

Cross validation is a widely known method to test the model prediction performance. In Bayesian statistics a similar technique was developed by Gelfand et al. (1990) based on the conditional predictive ordinate (CPO). For the i-th observation $y_i = \log t_i$, this is defined by $\text{CPO}_i = f(y_i|\boldsymbol{y}_{-i}) = \int f(y_i|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{y}_{-i})d\boldsymbol{\theta}$, where $\boldsymbol{y}_{-i} = \boldsymbol{y} \setminus \{y_i\}$. Then the log pseudo marginal likelihood (LPML) is constructed based on CPO as $\text{LPML} = \log \prod_{i=1}^{n} \text{CPO}_i$. By construction, a model with higher LPML is preferred. In a time to data analysis setting LPML has been routinely used, for example, see Ibrahim et al. (2002, 2005). When compared, the LPML for our proposed group-corr method is higher than when the regressions are fitted with in each kidney cancer group (Table S4). This implies that the proposed method provides a better fit to the data with respect to the cross validation technique such as LPML.

Table S4: LPML results for TCPA kidney cancer data.

Method	LPML
local	-3052.78
group-corr	-1494.08

Web Appendix F Tumors in Female Body

There exist at least 4 types of tumors – Breast invasive carcinoma (BRCA), Ovarian serous cystadenocarcinoma (OV), Uterine Corpus Endometrial Carcinoma (UCEC), and Uterine Carcinosarcoma (UCS), which are related to female body only. So the interest is to jointly analyze the protein data for these tumors and to find out the common proteins associated with these tumors. Table S5 provides the estimated figures of new cancer cases and deaths caused by these tumors in 2017 in the United States. Based on available TCPA data, in Figure S4, we plot the observed Kaplan-Meier plots of these tumors. The BRCA data has 871 samples, the OV data has 430 samples and the two Uterine cancers are consist of 436 and 48 samples respectively. All four groups have 189 proteins as before.

We apply our developed methodologies in this data analysis to recover the major proteins causing the cancers in female body. After running 4 MCMC chains the mean IBS produced by group-corr method and all-corr method coincides at 0.170 while the same due to the local method is 0.430 indicating the better predictive ability of the correlation structures. For simplicity, again, we carry out the following analyses for group-corr method only because both correlation structures have same IBS.

Table S5: Estimated number of new Cancer Cases and Deaths in US, 2017 (American Cancer Society 2017 report).

Tumor	Cases	Deaths
Breast invasive carcinoma	252,710	40,610
Ovarian serous cystadenocarcinoma	22,440	14,080
Uterine Corpus	61,380	10,920

In Figure S5, we depict the posterior estimates of the protein effects for local and group-corr methods respectively. Next we identify top 14 significant proteins which cause the cancer in a female body using the method described in Section 3.1 in the main article. We run 4 MCMC chains and the selected proteins for all 4 chains are listed in Table S6. Note that, some of the proteins (e.g. FOXO3A_pS318S321, DIRAS3, SF2) are also significant related to kidney cancers. Nevertheless, Taylor et al. (2015) provided a detailed review how FOXO3A has being targeted for breast cancer therapeutics. Moreover, Levanon et al. (2014) suggested that the same FOXO3A should be targeted for ovary cancer therapeutics. Similar conclusions have been drawn by Myatt et al. (2010) for uterine cancers. Among other selected proteins the possible effect of RAD51 has been discussed in Lose et al. (2006), in Hu and Sun (2015), and in Thacker (2005) for breast cancer, for ovary cancer, and for uterine cancers respectively. The association of other proteins for development and progression of female cancers are also well studied in the literature.

Table S6: Selected top 14 proteins for female cancers for 4 MCMC chains.

Chain 1	Chain 2	Chain 3	Chain 4
DIRAS3	DIRAS3	DIRAS3	DIRAS3
FOXO3A_pS318S321	SF2	RAD51	FOXO3A_pS318S321
PI3KP85	SHC_pY317	DIRAS3	SF2
BAK	BAK	PI3KP85	SHC_pY317
RAD51	FOXO3A_pS318S321	FOXO3A_pS318S321	MTOR
SF2	PI3KP85	SHC_pY317	BAK
MTOR	MTOR	MTOR	MSH2
PCADHERIN	RAD51	BAK	BCLXL
DIRAS3	PRDX1	PCADHERIN	PEA15
RAPTOR	PAI1	RAPTOR	RAD51
SHC_pY317	CYCLINE2	BAX	DIRAS3
MSH2	DIRAS3	SF2	KU80
CD31	RBM15	PAI1	SHC_pY317
BCLXL	SF2	BCLXL	PI3KP85
EGFR_pY1068	BCLXL	KU80	SMAD4

To confirm that the selected proteins are indeed important for the survival of the subjects we carry out the following analysis. First, after fitting a log normal AFT model using these protein expressions the IBS is computed which is 0.187 which is close to the IBS obtained

by group-corr method. We also randomly select 14 proteins and run a log normal AFT model using these proteins. After repeating this procedure the resulting mean IBS of these models is 0.669 which is higher than 0.187 and higher than the original IBS 0.170 obtained by group-corr method.

Web Appendix G Correlation Between Coefficients of Two Groups

In the following we derive the correlation between two tumor groups k and k', $k \neq k'$, given the prior specification (1).

$$\begin{aligned}\text{Var}(\beta_{jk}) =& E\{\text{Var}(\beta_{jk}|b_{P_j})\} + Var\{E(\beta_{jk}|b_{P_j})\} \\ =& E(\lambda_{jk}^2\tau^2\sigma^2) + \text{Var}(b_{P_j}) \\ =& \lambda_{jk}^2\tau^2\sigma^2 + \sigma_P^2.\end{aligned}$$

Similarly, $\text{Var}(\beta_{jk'}) = \lambda_{jk'}^2\tau^2\sigma^2 + \sigma_P^2$.

$$\begin{aligned}\text{Cov}(\beta_{jk}, \beta_{jk'}) =& E\{\text{Cov}(\beta_{jk}, \beta_{jk'}|b_{P_j})\} + \text{Cov}\{E(\beta_{jk}|b_{P_j}), E(\beta_{jk'}|b_{P_j})\} \\ =& 0 + \text{Cov}(b_{P_j}, b_{P_j}) \\ =& \text{Var}(b_{P_j}) \\ =& \sigma_P^2.\end{aligned}$$

This follows that, $\text{Corr}(\beta_{jk}, \beta_{jk'}) = \sigma_P^2/\{(\lambda_{jk}^2\tau^2\sigma^2 + \sigma_P^2)^{1/2}(\lambda_{jk'}^2\tau^2\sigma^2 + \sigma_P^2)^{1/2}\}$.

Web Appendix H Correlation Between Coefficients of Two Proteins

While borrowing strength among groups can be achieved using the prior structure as discussed in Section 2.2, one can argue for similar assumption of correlations among proteins that is the proteins are correlated for each individual. This can be accomplished by simple addition of a mean parameter in prior (1), $\beta_{jk}|\lambda_{jk}, \tau, \sigma^2 \sim N(b_{Pj} + b_{Gk}, \lambda_{jk}^2\tau^2\sigma^2)$, and $b_{Gk} \sim N(0, \sigma_G^2)$.

Here we derive the correlation induced by the prior specification in Section 2.3. We note that, $\text{Cov}(\beta_{jk}, \beta_{j'k}) = \sigma_G^2$ and $\text{Var}(\beta_{jk}) = \lambda_{jk}^2\tau^2\sigma^2 + \sigma_P^2 + \sigma_G^2$. This follows that, $\text{Corr}(\beta_{jk}, \beta_{j'k}) = \sigma_G^2/\{(\lambda_{jk}^2\tau^2\sigma^2 + \sigma_P^2 + \sigma_G^2)^{1/2}(\lambda_{j'k}^2\tau^2\sigma^2 + \sigma_P^2)^{1/2} + \sigma_G^2\}$.

Web Appendix I Additional Simulations

In this section we consider additional simulation studies by keeping the same settings as in Section for except set the effect size as 0.2 or -0.2 randomly. The area under the ROC curve (AUC) results are given in Table S7, when we assume the independence among the groups.

Table S7: Area under the ROC curves when groups are independent

Censoring Rate	35%	48%	76%
local	0.659	0.646	0.562
group-corr	0.690	0.668	0.577
all-corr	0.688	0.662	0.565
lasso	0.419	0.412	0.449

Table S8: Area under the ROC curves when groups are correlated.

Censoring Rate	35%	56%	76%
local	0.645	0.641	0.488
group-corr	0.699	0.680	0.586
all-corr	0.686	0.677	0.574
lasso	0.487	0.498	0.512

Furthermore, when we consider a correlation structure among groups then the results are provided in Table S8. We notice that the results are consistent with those found in Section 4 i.e. the group-corr method continues to be superior among the methods considered.

Web Appendix J Integrated Brier Score

One way to measure the prediction accuracy is to plot the observed Kaplan-Meier curve along with the Kaplan-Meier curve based on samples from the posterior predictive model (see Banerjee et al. 2003). We take a step further to calculate Brier score (BS) introduced by Graf et al. (1999), $\text{BS}(t) = n^{-1} \sum_{i=1}^{n} \left[\frac{\widehat{S}(t|x_i)^2 I(t_i \leq t \wedge \delta_i = 1)}{\widehat{K}(t_i)} + \frac{(1-\widehat{S}(1|x_i))^2 I(t_i > t)}{\widehat{K}(t)} \right]$, where $\widehat{K}(.)$ denotes the Kaplan-Meier estimate of the censoring distribution which is based on the observations $(t_i, 1-\delta_i)$, and $\widehat{S}(.)$ stands for the estimated survival function. As the mathematical form suggests, BS provides a numerical comparison between observed and estimated survival functions. It has been shown useful to measure the goodness of fit of a survival model (Hothorn et al., 2006; Schumacher et al., 2007; Bonato et al., 2011). BS is defined for each time point t, and hence can be added for the entire time range to obtain Integrated Brier score, $\text{IBS} = \max(t_i)^{-1} \int_0^{\max(t_i)} \text{BS}(t) dt$. We can see that, models with smaller scores are preferred. Following Van Wieringen et al. (2009), we compute IBS using **ipred** package.

References

Analytics, R. and Weston, S. (2015a). *doParallel: Foreach Parallel Adaptor for the 'parallel' Package*. R package version 1.0.10.

Analytics, R. and Weston, S. (2015b). *foreach: Provides Foreach Looping Construct for R*. R package version 1.4.3.

Armagan, A., Dunson, D. B., Lee, J., Bajwa, W. U., and Strawn, N. (2013). Posterior consistency in linear models under shrinkage priors. *Biometrika*, 100(4):1011–1018.

Banerjee, S., Wall, M. M., and Carlin, B. P. (2003). Frailty modeling for spatially correlated survival data, with application to infant mortality in Minnesota. *Biostatistics*, 4(1):123–142.

Bates, D. and Maechler, M. (2016). *Matrix: Sparse and Dense Matrix Classes and Methods*. R package version 1.2-7.1.

Bonato, V., Baladandayuthapani, V., Broom, B. M., Sulman, E. P., Aldape, K. D., and Do, K.-A. (2011). Bayesian ensemble methods for survival prediction in gene expression data. *Bioinformatics*, 27(3):359–367.

Bradic, J., Fan, J., and Jiang, J. (2011). Regularization for Cox's proportional hazards model with NP-dimensionality. *Annals of Statistics*, 39(6):3092.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1.

Gelfand, A. E., Hills, S. E., Racine-Poon, A., and Smith, A. F. (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association*, 85(412):972–985.

Genz, A. and Bretz, F. (2009). *Computation of multivariate normal and t probabilities*, volume 195. Springer Science & Business Media.

Graf, E., Schmoor, C., Sauerbrei, W., and Schumacher, M. (1999). Assessment and comparison of prognostic classification schemes for survival data. *Statistics in Medicine*, 18(17-18):2529–2545.

Hothorn, T., Bühlmann, P., Dudoit, S., Molinaro, A., and Van Der Laan, M. J. (2006). Survival ensembles. *Biostatistics*, 7(3):355–373.

Hu, X. and Sun, S. (2015). RAD51 Gene 135G/C polymorphism and ovarian cancer risk: a meta-analysis. *International journal of clinical and experimental medicine*, 8(12):22365.

Ibrahim, J. G., Chen, M.-H., and Gray, R. J. (2002). Bayesian models for gene expression with DNA microarray data. *Journal of the American Statistical Association*, 97(457):88–99.

Ibrahim, J. G., Chen, M.-H., and Sinha, D. (2005). *Bayesian Survival Analysis*. Wiley Online Library.

Jackson, C. H. (2011). Multi-State Models for Panel Data: The msm Package for R. *Journal of Statistical Software*, 38(8):1–29.

Levanon, K., Sapoznik, S., Bahar-Shany, K., Brand, H., Shapira-Frommer, R., Korach, J., Hirsch, M. S., Roh, M. H., Miron, A., Liu, J. F., et al. (2014). FOXO3a loss is a frequent early event in high-grade pelvic serous carcinogenesis. *Oncogene*, 33(35):4424.

Lose, F., Lovelock, P., Chenevix-Trench, G., Mann, G. J., Pupo, G. M., and Spurdle, A. B. (2006). Variation in the RAD51 gene and familial breast cancer. *Breast Cancer Research*, 8(3):R26.

Myatt, S. S., Wang, J., Monteiro, L. J., Christian, M., Ho, K.-K., Fusi, L., Dina, R. E., Brosens, J. J., Ghaem-Maghami, S., and Lam, E. W.-F. (2010). Repression of FOXO1 expression by microRNAs in endometrial cancer. *Cancer research*, 70(1):367.

Neal, R. M. (2003). Slice sampling. *Annals of Statistics*, pages 705–741.

Peters, A. and Hothorn, T. (2017). *ipred: Improved Predictors*. R package version 0.9-6.

Polson, N. G., Scott, J. G., and Windle, J. (2014). The Bayesian bridge. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(4):713–733.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., and Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, 12:77.

Schumacher, M., Binder, H., and Gerds, T. (2007). Assessment of survival prediction models based on microarray data. *Bioinformatics*, 23(14):1768–1774.

Taylor, S., Lam, M., Pararasa, C., Brown, J. E., Carmichael, A. R., and Griffiths, H. R. (2015). Evaluating the evidence for targeting FOXO3a in breast cancer: a systematic review. *Cancer cell international*, 15(1):1.

Thacker, J. (2005). The RAD51 gene family, genetic instability and cancer. *Cancer letters*, 219(2):125–135.

Van Wieringen, W. N., Kun, D., Hampel, R., and Boulesteix, A.-L. (2009). Survival prediction using gene expression data: a review and comparison. *Computational Statistics & Data Analysis*, 53(5):1590–1603.

Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wu, Y. and Ghosal, S. (2008). Posterior consistency for some semi-parametric problems. *Sankhyā: The Indian Journal of Statistics, Series A (2008-)*, 70(3):267–313.

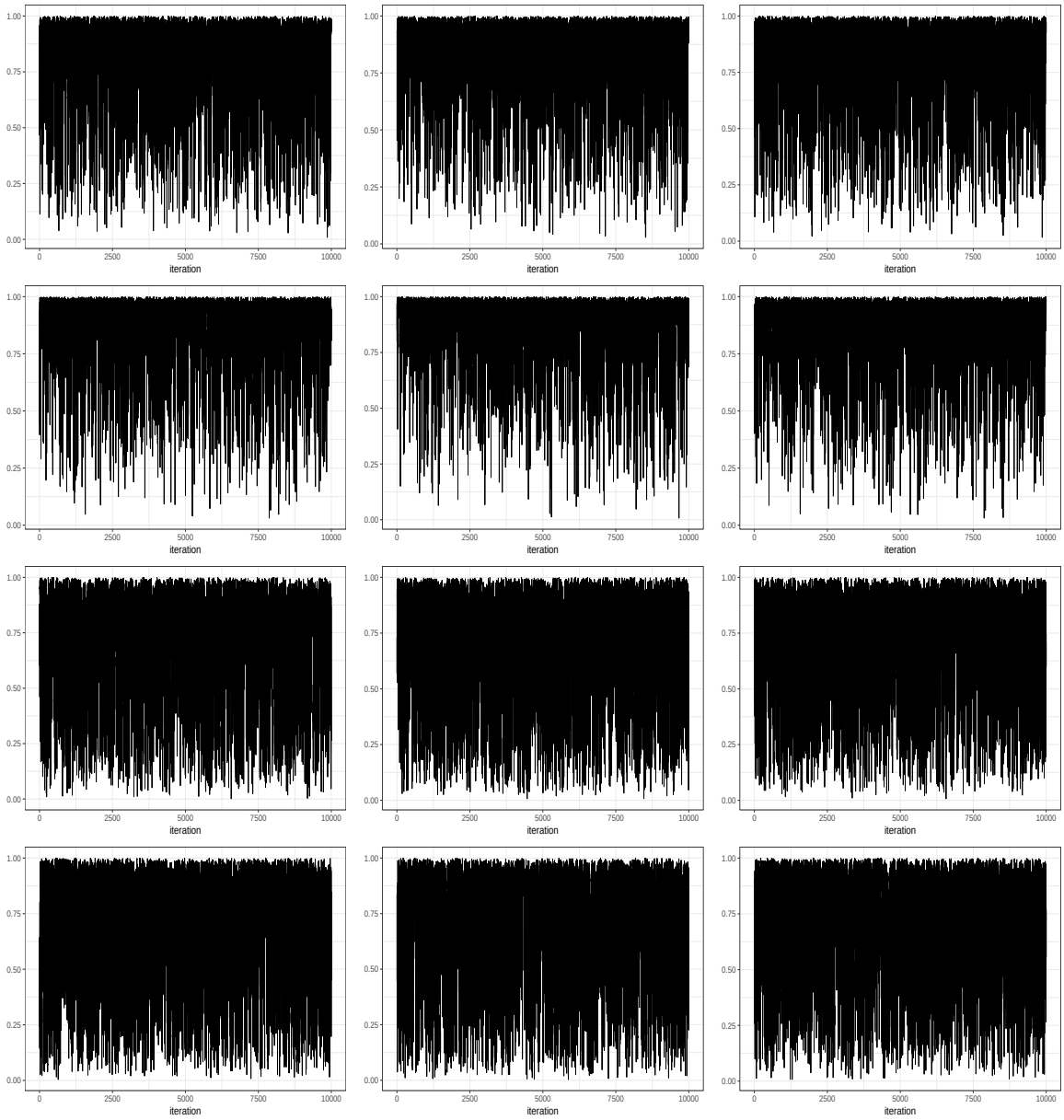

Figure S2: Trace plots of posterior correlations. The top panel shows the trace plots of correlations between KIRC and KIRP, KICH and KIRC, and KICH and KIRP kidney tumor groups respectively for X1433EPSILON protein. Similarly, the second, third, and the fourth panels plots correlations for X4EBP1, FOXO3A_pS318S321, and DIRAS3 proteins respectively.

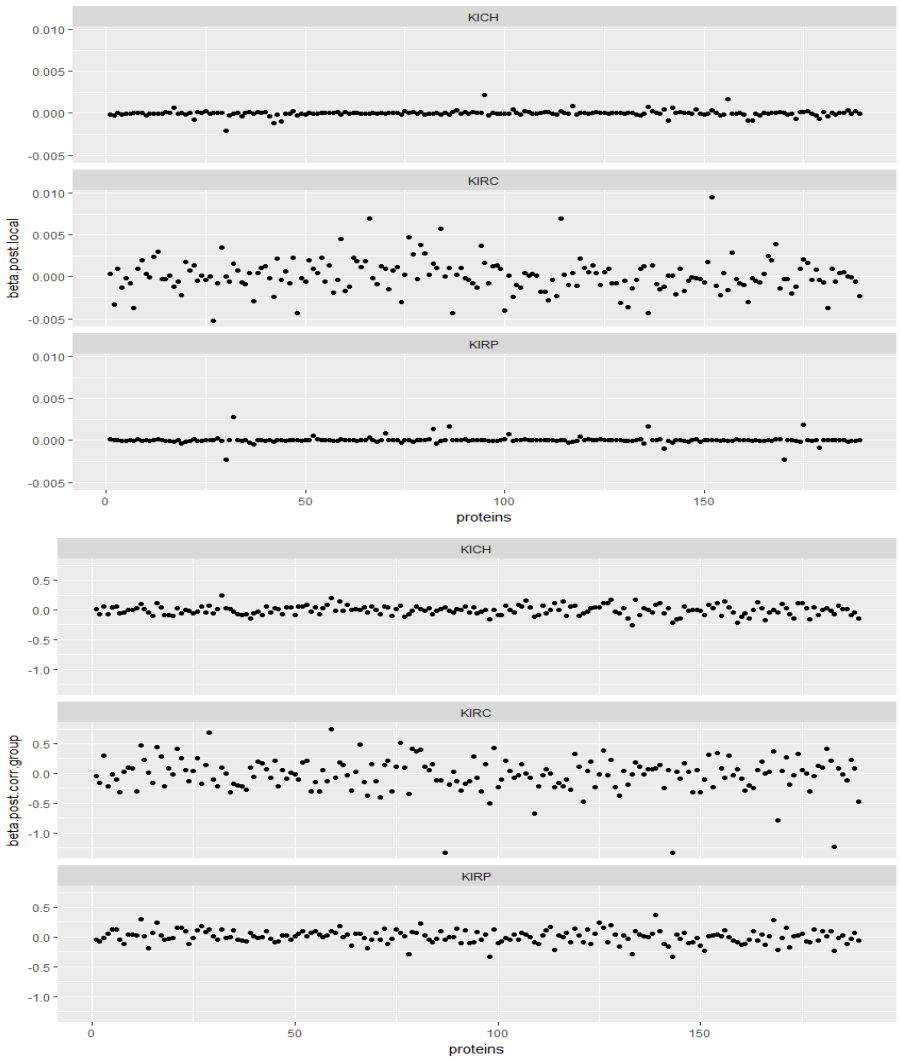

Figure S3: The upper panel plot is the posterior estimates of protein effects for different tumor groups when regressions are run for each group separately in Kidney cancer data. The lower panel plot is same for our proposed group-corr method.

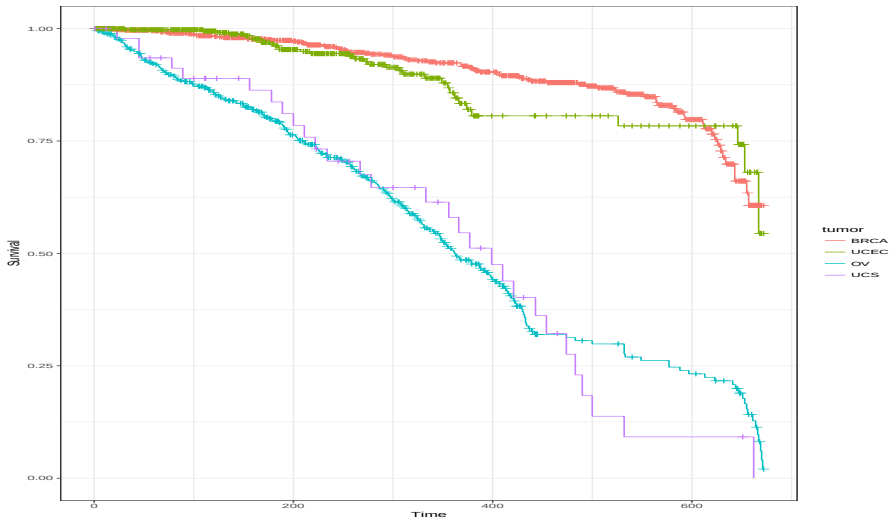

Figure S4: Observed Kaplan-Meier plots for female tumors – BRCA, OV, UCEC, and UCS.

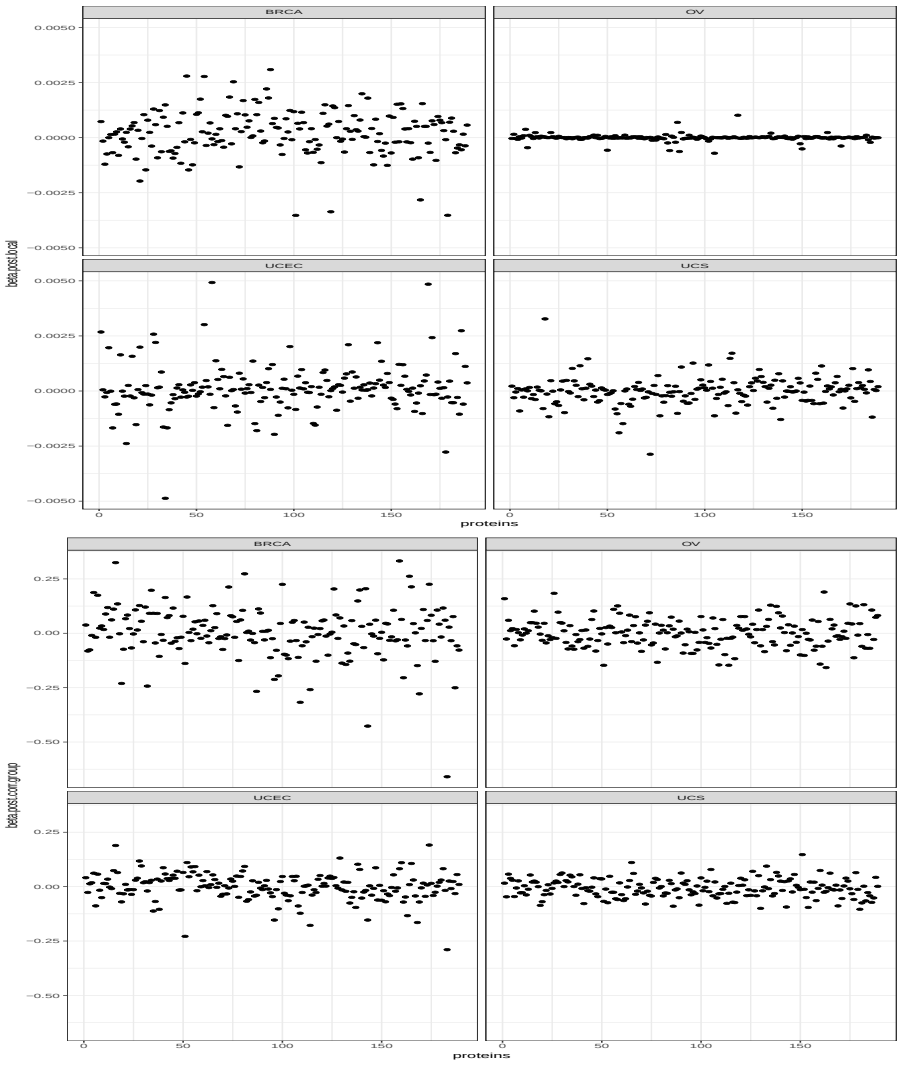

Figure S5: The upper panel plot is the posterior estimates of protein effects for different female tumor groups when regressions are run for each group separately for female body tumors. The lower panel plot is same for our proposed group-corr method.