Journal of Statistical Research *https://doi.org/10.3329/jsr.v56i1.63947*
2022, Vol. 56, No. 1, pp. 75-100 *ISSN 0256 - 422 X*

PROBABILISTIC CANONICAL CORRELATION ANALYSIS FOR SPARSE COUNT DATA

LIN QIU
The Pennsylvania State University, State College, PA, USA
Email: lin.qiu.stats@gmail.com

VERNON M. CHINCHILLI*
The Pennsylvania State University, Hershey, PA, USA
Email: vmc1@psu.edu

SUMMARY

Canonical correlation analysis (CCA) is a classical and important multivariate technique for exploring the relationship between two sets of continuous variables. CCA has applications in many fields, such as genomics and neuroimaging. It can extract meaningful features as well as use these features for subsequent analysis. Although some sparse CCA methods have been developed to deal with high-dimensional problems, they are designed specifically for continuous data and do not consider the integer-valued data from next-generation sequencing platforms that exhibit very low counts for some important features. We propose a model-based probabilistic approach for correlation and canonical correlation estimation for two sparse count data sets. Probabilistic sparse CCA (PSCCA) demonstrates that correlations and canonical correlations estimated at the natural parameter level are more appropriate than traditional estimation methods applied to the raw data. We demonstrate through simulation studies that PSCCA outperforms other standard correlation approaches and sparse CCA approaches in estimating the true correlations and canonical correlations at the natural parameter level. We further apply the PSCCA method to study the association of miRNA and mRNA expression data sets from a squamous cell lung cancer study, finding that PSCCA can uncover a large number of strongly correlated pairs than standard correlation and other sparse CCA approaches.

Keywords and phrases: Canonical correlation analysis (CCA); Sparse count data; High-dimension

1 Introduction

Recent advancements in next-generation sequencing (NGS) technology have enabled the measurement of multiple high-dimensional data types in a single study, such as genomics, transcriptomics, epigenomics, and metabolomics. Integrative analysis of high-dimensional omics data is becoming increasingly important and popular. It has been shown that combining multiple omics data types can

* Corresponding author

improve analysis and lead to biologically more meaningful results for complex diseases (Safo et al., 2018; Lu et al., 2019).

Sequencing-based omics datasets have three main characteristics that pose modeling challenges: (1) they are high-dimensional data, with a large number of variables p and small sample size n; (2) the raw data represent count variables which violate the distributional assumptions for standard correlation and canonical correlation analysis, which can lead to invalid inference in the presence of a small sample size; (3) the data are very sparse, with a large proportion of these counts being very close to zero and having random missing values.

1.1 CCA

Canonical correlation analysis (CCA) is a classical multivariate method proposed by Hotelling (1936) for exploring the relationship between two sets of variables. Consider two random vectors $\mathbf{X} \in \mathbb{R}^p$, $\mathbf{Y} \in \mathbb{R}^q$. Define $\sum_{\mathbf{XX}}$ = cov(**X**), $\sum_{\mathbf{YY}}$ = cov(**Y**), and $\sum_{\mathbf{XY}}$ = cov(**X**,**Y**). CCA seeks vectors $\mathbf{a} \in \mathbb{R}^p$ and $\mathbf{b} \in \mathbb{R}^q$ that solve

$$\underset{\mathbf{a},\mathbf{b}}{\operatorname{argmax}} \operatorname{corr}\left(\mathbf{a}^\top \mathbf{X}, \mathbf{b}^\top \mathbf{Y}\right)$$

subject to $\mathbf{a}^{\mathbf{T}} \sum_{\mathbf{XX}} \mathbf{a}$=1, $\mathbf{b}^{\mathbf{T}} \sum_{\mathbf{YY}} \mathbf{b}$=1, then the optimization can be attained by applying the singular value decomposition (SVD) and replacing $\sum_{\mathbf{XX}}^{-1/2} \sum_{\mathbf{XY}} \sum_{\mathbf{YY}}^{-1/2}$ with their sample estimates $\hat{\sum}_{\mathbf{XX}}^{-1/2} \hat{\sum}_{\mathbf{XY}} \hat{\sum}_{\mathbf{YY}}^{-1/2}$. However, in a high-dimensional setting, when the dimensions $p, q \gg$ n, the SVD approach is not applicable because $\hat{\sum}_{\mathbf{XX}}$ and $\hat{\sum}_{\mathbf{YY}}$ are not invertible.

1.2 Related work

Motivated by genomics, neuroimaging and other applications, researchers have been working on generalizing CCA to accommodate high dimensions, usually called sparse CCA (Witten and Tibshirani, 2009; Avants et al., 2010; Hardoon and Shawe-Taylor, 2011; Gao et al., 2017). These methods impose sparsity constraints on the canonical directions which effectively can reduce the dimensionality and improve the interpretation of the correlations. Penalized matrix decomposition (PMD) (Witten and Tibshirani, 2009) is one of the most popular sparse CCA methods, which uses the penalized matrix decomposition to replace $\hat{\sum}_{\mathbf{XX}}$ and $\hat{\sum}_{\mathbf{YY}}$ with identity matrices to avoid singularities. By doing so, PMD can obtain sparse estimates of the canonical directions by penalization. However, PMD may perform poorly on data sets when $\sum_{\mathbf{XX}}$ and $\sum_{\mathbf{YY}}$ are far from diagonal. With respect to genomics data, for example, genes usually have strong correlations among them.

The probabilistic interpretation of CCA was initiated by Bach and Jordan (2006). Later on, several Bayesian versions of CCA were developed (Archambeau and Bach, 2008; Virtanen et al., 2011; Klami et al., 2013). One of the key promising features of Bayesian CCA is that it enables analysis of high-dimensional data in life sciences (Fujiwara et al., 2009; Huopaniemi et al., 2010). However, these methods assume the data to follow normal distributions. Thus, the aforementioned Bayesian methods may not work well for non-normally distributed data. PCAN is the first approach that describes a Bayesian correlation analysis method for count data (Zoh et al., 2016), in which

the correlations are estimated based on the latent weights from the natural parameters of the data generating model, rather than the correlations among the counts. In the latent variable model, priors or strong regularizations are used on the latent weights to induce sparsity (West, 2003).

The rise of big datasets with few signals, such as omics datasets, has spurred the study of sparse models. From a Bayesian perspective, discrete mixtures (George and McCulloch, 1993) and shrinkage priors (Tipping, 2001) are the two main sparse estimation methods. In latent variable models, Bayesian shrinkage priors are popular because of their flexible and interpretable solutions (Carvalho et al., 2008; Knowles and Ghahramani, 2011; Bhattacharya and Dunson, 2011). The spike-and-slab prior is a mixture of a point mass at zero and a flat distribution across the space of real values. The excluded loadings are modeled by the "spike" distribution, whereas the included loadings are modeled by the "slab" distribution (Carvalho et al., 2008). Structural Bayesian prior is proposed to encourage both element-wise and column-wise shrinkage and leads to desirable behavior on high-dimensional data (Zhao et al., 2016). The disadvantages of these models are that the results can be sensitive to prior choices and it is computationally demanding for posterior inference with a large number of variables due to a huge model space. Scale mixtures of normal priors have been proposed recently as a computationally efficient alternative to the two component spike-and-slab prior (Armagan et al., 2013; Bhattacharya et al., 2014).

These types of priors usually assume normal distributions with a mixed variance term and the mixing variance distribution enables strong shrinkage close to zero. For example, Bayesian canonical correlation analysis (BCCA) (Klami et al., 2013) consists of applying an automatic relevance determination (ARD) (Neal, 1996) prior for the latent weights which is a normal-gamma prior that imposes an inverse gamma distribution on the variance term. The horseshoe prior is popular due to its good performance in simulations and under theoretical study, which has shown comparable performance to the spike-and-slab prior in a variety of problems where a sparse prior is desirable (Carvalho et al., 2008, 2010; Polson and Scott, 2011).The horseshoe prior is a scale mixture of normals, with a product of half-Cauchy priors on the variance. It is given by

$$\begin{aligned} \theta_i|\lambda_i, \tau &\sim N(0, \lambda_i^2\tau^2), \\ \lambda_i &\sim C^+(0,1),\ i = 1, \ldots, n. \end{aligned}$$

The global hyperparameter τ can shrink all the parameters toward zero, especially if its domain is restricted to a finite interval, while the heavy-tailed half-Cauchy local priors allow some parameters to escape. Different levels of sparsity can be accommodated by changing the value of τ: the large τ will have little shrinkage, while small τ will shrink all the weights to zero. Despite the good performance, there are two shortcomings for the horseshoe prior. First, how to perform inference for the global hyperparameter τ which determines the overall sparsity in the parameter vector $\boldsymbol{\theta}$ is not fully answered yet. Second, parameters far from zero will not be regularized at all. Quite a few researchers have investigated the impact of τ concerning the resulting posterior distribution both for recovery and for uncertainty quantification, either in a deterministic way or a hierarchical full Bayes approach (Carvalho et al., 2008; Datta and Ghosh, 2013; Pas et al., 2014, 2017). We take the second shortcoming as the key strength of this prior and to incorporate it with a latent variable model to infer the feature sparsity jointly. For an omics data set we assume only important variables

are strongly identified and the parameters far from zero will not be regularized.

1.3 Our contribution

In this study, we propose a new probabilistic framework of CCA for sparse count data, which we label a probabilistic sparse canonical correlation analysis (PSCCA). Our work contributes several important advances. First, we propose to estimate the canonical correlations at the natural parameter level for data expressed as raw counts, which is lacking in sequencing-based omics analyses. Second, we provide a theoretical justification for estimating the correlations and canonical correlations based on the natural parameters rather than based on the raw data. The former are larger in magnitude than the latter, which is very meaningful for CCA. Because CCA is an exploratory analytical method, larger values of the canonical correlations yield less chance to miss the true correlation pairs. Third, the horseshoe prior is widely studied in the literature, via both simulation studies and theoretical research. Nevertheless, we do not see many examples in applications. We formulate the natural parameters as a latent variable model, and we invoke the horseshoe prior for the latent weight to model the sparsity. To better extract the sparse signals we assume $\tau \sim C^{+}(0,1)$ for the global hyperparameter. As discussed in Piironen and Vehtari (2017), this prior results in sensible inference only when τ is strongly identified by the data. Our simulation study and real data applications show that our approach performs better than existing methods. Lastly, our approach is built on an exponential family and can be easily extended to other formats of data.

The rest of the article is organized as follows. Section 2 contains our model details and inference. Section 3 discusses the theoretical results and Section 4 describes simulation studies. Section 5 presents the real data application. Finally, Section 6 contains a discussion and future directions of PSCCA.

2 Method

2.1 Model

Let $\mathrm{F}_y(\cdot|\cdot)$ be a distribution function from the natural parameter exponential family. The random component of a generalized linear model consists of a response vector $\mathbf{y} \in \mathbb{R}^N$ which has a conditional distribution in the exponential family. This family has probability density function or mass function of the form $f_y(y_j, \theta_j) = a(\theta_j)b(y_j)\exp[y_j\mathcal{Q}(\theta_j)]$. The value of the parameter θ_j may vary for $j = 1, \ldots, N$ depending on values of the explanatory variables. The term $\mathcal{Q}(\theta)$ is the natural parameter; $a(\cdot)$ and $b(\cdot)$ are non-negative functions that distinguish one member of the exponential family from another. For our case, assume we have two sets of multivariate random variables, $\boldsymbol{Y}^{(1)} \in \mathbb{R}^{D_1 \times 1}$, and $\boldsymbol{Y}^{(2)} \in \mathbb{R}^{D_2 \times 1}$. The observed data samples are expressed as $[\boldsymbol{Y}_1^{(m)}, \ldots, \boldsymbol{Y}_N^{(m)}] \in \mathbb{R}^{D_m \times N}$ with N observations, where m is 1 or 2. Let $y_{ij}^{(m)}$ represent the observed value of the j^{th} individual for the i^{th} feature (variable) in a set of D_m measured features (variables).

We motivate our formulation in the latent variable interpretation of CCA (Bach and Jordan, 2006) to model the natural parameters and the ideas from PCAN (probabilistic correlation analysis)

(Zoh et al., 2016) for the correlation estimation from the natural parameters. We assume each individual data vector follows conditionally an exponential family distribution and here we consider a generalized linear model. The generative model for D_m coupled natural parameters $\boldsymbol{\theta}_{.j}^{(m)}$ with m=1, 2 and $j = 1, \ldots, N$ is

$$\boldsymbol{\theta}_{.j}^{(m)} = \boldsymbol{\mu}_{\theta}^{(m)} + \mathbf{W}^{(m)}\mathbf{Z}_j + \boldsymbol{\epsilon}_{.j}^{(m)}. \tag{2.1}$$

The parameter vector $\boldsymbol{\mu}_{\theta}^{(m)}$ represents the mean of the natural parameters associated with the D_m features in the vectors $\boldsymbol{Y}_{.j}^{(m)}$; the matrix $\mathbf{W}^{(m)} \in \mathbb{R}^{D_m \times d}$ denotes the loading matrix associated with the latent vector $\mathbf{Z}_j = (Z_{1j}, \ldots, Z_{dj})^T$; and $\boldsymbol{\epsilon}_{.j}^{(m)}$ is an independently distributed random error vector.

We write the model as a function of latent variables and the D_m features of the vector $\boldsymbol{\theta}_{.j}^{(m)}$ as follows

$$\begin{aligned} &\boldsymbol{Y}_{.j}^{(m)} | \boldsymbol{\theta}_{.j}^{(m)} \sim \text{Poisson}\left\{\exp(\boldsymbol{\theta}_{.j}^{(m)})\right\}, \\ &\boldsymbol{\epsilon}_{.j}^{(m)} \sim N_{D_m}(\mathbf{0}_{D_m \times D_m}, {\sigma_{\theta}^{(m)}}^2 \boldsymbol{I}_{D_m \times D_m}), \\ &\mathbf{Z}_j \sim N_d(\mathbf{0}, \boldsymbol{I}_{d \times d}). \end{aligned} \tag{2.2}$$

The core generative process is the unobserved shared latent variables $\mathbf{Z}_j$, which are transformed via linear mappings to the observation spaces, and can capture the variation common to both data sets and allow for dependency between variables in a specific data set.

We impose horseshoe priors on the $D_m \times d$ matrix $\mathbf{W}^{(m)}$, and we let $\boldsymbol{W}_{i.}^{(m)}$ denote the ith row vector of $\boldsymbol{W}^{(m)}$. Then we assume that

$$\boldsymbol{W}_{i.}^{(m)} | \lambda_i^{(m)}, \tau^{(m)} \sim \mathrm{N}(\mathbf{0}, {\lambda_i^{(m)}}^2 {\tau^{(m)}}^2 \boldsymbol{I}_{d \times d}). \tag{2.3}$$

We refer to the $\lambda_i^{(m)}$ as the local shrinkage parameters and to $\tau^{(m)}$ as the global shrinkage parameters. Let $C^+(0, 1)$ denote the standard half-Cauchy distribution. The half-Cauchy prior for the local shrinkage parameter λ_i has shown good performance (Carvalho et al., 2008, 2010). There has been a vast amount of research on how to choose the prior for the global hyperparameter τ which plays an important role in the overall sparsity for the parameter matrix $\boldsymbol{W}^{(m)}$. As discussed in section 1.2, we choose the full Bayesian specification for τ. Thus, we assume

$$\lambda_i^{(m)} \sim C^+(0, 1); \quad \tau^{(m)} \sim C^+(0, 1). \tag{2.4}$$

For $i = 1, \ldots, D_1$; $k = 1, \ldots, D_2$; and $j = 1, \ldots, N$, we construct $\boldsymbol{\theta}_{.j}^{(1)} = (\theta_{1j}^{(1)}, \ldots, \theta_{D_1 j}^{(1)})^T$ and $\boldsymbol{\theta}_{.j}^{(2)} = (\theta_{1j}^{(2)}, \ldots, \theta_{D_2 j}^{(m)})^T$. The vector $(\boldsymbol{\theta}_{.j}^{(1)} \quad \boldsymbol{\theta}_{.j}^{(2)})^T$ has a conditional multivariate normal distribution with unconditional mean $\boldsymbol{\mu}_{\theta} = (\mu_{\theta_1}^{(1)}, \ldots, \mu_{\theta_{D_1}}^{(1)}, \mu_{\theta_1}^{(2)}, \ldots, \mu_{\theta_{D_2}}^{(2)})^T$ and conditional covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} \mathbf{W}^{(1)}{\mathbf{W}^{(1)}}^T + {\sigma_{\theta}^{(1)}}^2 \mathbf{I_1} & \mathbf{W}^{(1)}{\mathbf{W}^{(2)}}^T \\ \mathbf{W}^{(2)}{\mathbf{W}^{(1)}}^T & \mathbf{W}^{(2)}{\mathbf{W}^{(2)}}^T + {\sigma_{\theta}^{(2)}}^2 \mathbf{I_2} \end{pmatrix}, \tag{2.5}$$

where $\mathbf{I_1} = \mathbf{I_{D_1 \times D_1}}$ and $\mathbf{I_2} = \mathbf{I_{D_2 \times D_2}}$. The conditional correlation between $\theta_{ij}^{(1)}$ and $\theta_{kj}^{(2)}$, for any sample j, can be obtained from

$$\begin{aligned} &\text{corr}(\boldsymbol{\theta}_{.j}^{(1)}, \boldsymbol{\theta}_{.j}^{(2)}) \\ &= diag^{-1/2}\big(\mathbf{W}^{(1)}\mathbf{W}^{(1)^T} + {\sigma_\theta^{(1)}}^2\mathbf{I_1}\big)\mathbf{W}^{(1)}\mathbf{W}^{(2)^T} diag^{-1/2}\big(\mathbf{W}^{(2)}\mathbf{W}^{(2)^T} + {\sigma_\theta^{(2)}}^2\mathbf{I_2}\big). \end{aligned} \tag{2.6}$$

For the canonical correlations, let $\boldsymbol{C}^{-1/2}$ denote the square-root decomposition of the positive definite matrix $\boldsymbol{C}^{-1}$. Let $\mathbf{R} = \big(\mathbf{W}^{(1)}\mathbf{W}^{(1)^T} + {\sigma_\theta^{(1)}}^2\mathbf{I_1}\big)^{-1}\mathbf{W}^{(1)}\mathbf{W}^{(2)^T}\big(\mathbf{W}^{(1)}\mathbf{W}^{(2)^T} + {\sigma_\theta^{(2)}}^2\mathbf{I_2}\big)^{-1}\mathbf{W}^{(2)}\mathbf{W}^{(1)^T}$, then the nonnull eigenvalues of $\mathbf{R}$ correspond to the squared canonical coefficients for the natural parameters. Inference on the correlations and canonical correlations will be based on the marginal posterior distribution of $\mathbf{W}^{(1)}$, $\mathbf{W}^{(2)}$, ${\sigma_\theta^{(1)}}^2$, and ${\sigma_\theta^{(2)}}^2$.

2.2 Identifiability and prior

The latent variable model (Equation 2.1) is identifiable up to orthonormal rotations, for any invertible $\mathbf{G} \in \mathcal{R}^{d\times d}$ with $\mathbf{G}^T\mathbf{G} = \mathbf{I}_{d\times d}$ (Muirhead, 1982). Then $\mathbf{W}^* = \mathbf{W}\mathbf{G}^T$ and $\mathbf{Z}^* = \mathbf{G}\mathbf{Z}$ will produce the same estimate of the conditional covariance matrix in Equation 2.5 and has an equal likelihood. Zoh et al. (2016) recommend imposing a lower triangular structure for the $d \times d$ upper submatrices of $\mathbf{W}^{(1)}$ and $\mathbf{W}^{(2)}$, following the work of (Geweke and Zhou, 1996), and they further require that the diagonal elements are non-negative to remove the non-identifiability related to the sign. This approach relies on the choice of d based on the D_m rows of $\mathbf{W}^{(m)}$, but we assume the value of d will be small so that the impact is negligible (Lopes and West, 2004), i.e.,

$$\begin{aligned} W_{ik}^{(m)} &\sim \text{N}(0, {\lambda_i^{(m)}}^2{\tau^{(m)}}^2) \quad \text{if} \quad i < k, \\ W_{ik}^{(m)} &\sim \text{N}(0, {\lambda_i^{(m)}}^2{\tau^{(m)}}^2)\mathbf{1}(W_{kk}^{(m)} > 0) \quad \text{if} \quad i = k, \end{aligned} \tag{2.7}$$

where for $m = 1$, $i = 1, \ldots, D_1, k = 1, \ldots, d$ and for $m = 2$, $i = 1, \ldots, D_2, k = 1, \ldots, d$. $\mathbf{1}$ is an indicator function, $\mathbf{1}$(W)=$\mathbf{1}$ if W is true and 0 otherwise. We assume conjugate priors for the remaining parameters in the model as

$$\begin{aligned} \boldsymbol{\mu}_\theta^{(m)} &\sim \Pi_{i=1}^{D_m}\text{Normal}(0, k_i^{(m)}), \\ {\sigma_\theta^{(m)}}^2 &\sim \text{Inv-}\chi^2(\nu_\theta^{(m)}, {s_\theta^{(m)}}^2), \end{aligned} \tag{2.8}$$

The hyperparameters $k_i^{(m)}$, ${s_\theta^{(m)}}^2$, and $\nu_\theta^{(m)}$ are determined by the analyst.

2.3 Inference

The form of the full conditional posterior distribution is proportional to the product of the joint conditional likelihood for the data matrices $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$ and the prior distributions

$$
\begin{aligned}
&P\Big(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \mathbf{W}^{(1)}, \mathbf{W}^{(2)}, \mathbf{Z}, {\sigma_\theta^{(1)}}^2, {\sigma_\theta^{(2)}}^2 | \mathbf{Y}^{(1)}, \mathbf{Y}^{(2)}\Big) \\
&\propto l\Big(\mathbf{Y}^{(1)}, \mathbf{Y}^{(2)} | \boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \mathbf{W}^{(1)}, \mathbf{W}^{(2)}, \mathbf{Z}, {\sigma_\theta^{(1)}}^2, {\sigma_\theta^{(2)}}^2\Big) \\
&\times \Pi_{i=1}^{d} \Bigg\{ (\lambda_i^{(1)} \tau^{(1)})^{D_1} \exp\big(-0.5 \lambda_i^{(1)} \tau^{(1)} \mathbf{W}_{i.}^{(1)} {\mathbf{W}_{i.}^{(1)}}^T \big) \\
&(\lambda_i^{(2)} \tau^{(2)})^{D_2} \exp\big(-0.5 \lambda_i^{(2)} \tau^{(2)} \mathbf{W}_{i.}^{(2)} {\mathbf{W}_{i.}^{(2)}}^T \big) \\
&\times \frac{1}{1+{\lambda_i^{(1)}}^2} \frac{1}{1+{\tau^{(1)}}^2} \frac{1}{1+{\lambda_i^{(2)}}^2} \frac{1}{1+{\tau^{(2)}}^2} \Bigg\} \\
&\times \Big\{ \Pi_{j=1}^{N} \exp(-0.5 \mathbf{Z_j^T Z_j}) \Big\} \exp\Big\{ -\nu_\theta^{(1)} {s_\theta^{(1)}}^2 / (2{\sigma_\theta^{(1)}}^2) \Big\} \\
&\times {\sigma_\theta^{(1)}}^{-2(1+\nu_\theta^{(1)}/2)} \exp\Big\{ -\nu_\theta^{(2)} {s_\theta^{(2)}}^2 / (2{\sigma_\theta^{(2)}}^2) \Big\} {\sigma_\theta^{(2)}}^{-2(1+\nu_\theta^{(2)}/2)}.
\end{aligned}
\tag{2.9}
$$

We update the parameters in a Markov chain Monte Carlo (MCMC). R code implementation exploits the package Rstan for fast computation, we release our code at https://github.com/lquvatexas/PSCCA. The details of priors are included in the Appendix.

3 Theoretical Results

Let $F(\cdot|\cdot)$ be a cumulative distribution function from the natural parameter exponential family as we discussed in the model section. We model

$$
Y_{ij}^{(m)} \sim F(Y|\theta_{ij}^{(m)}) = \text{Poisson}(y|\theta_{ij}^{(m)}), \tag{3.1}
$$

where $m = 1, 2$, $i = 1, 2, \ldots, D_m$, $j = 1, 2, \ldots, N$. Let $\boldsymbol{Y}_{.j}^{(m)} = [Y_{1j}^{(m)}, Y_{2j}^{(m)}, \ldots, Y_{D_m j}^{(m)}]^T$ and $\boldsymbol{\theta}_{.j}^{(m)} = [\theta_{1j}^{(m)}, \theta_{2j}^{(m)}, \ldots, \theta_{D_m j}^{(m)}]^T$. Then, as in Equations (2.1) and (2.2),

$$
\boldsymbol{\theta}_{.j} = \boldsymbol{\mu}_\theta^{(m)} + \boldsymbol{W}^{(m)} \boldsymbol{Z}_j + \boldsymbol{\epsilon}_{.j}^{(m)},
$$

where $\boldsymbol{W}^{(m)} \in \mathbb{R}^{D_m \times d}$, $\boldsymbol{Z}_j \sim N_d(\mathbf{0}_{d\times d}, \boldsymbol{I}_{d\times d})$ and $\boldsymbol{\epsilon}_{.j}^{(m)} \sim N_{D_m}(\mathbf{0}_{D_m \times D_m}, {\sigma_\theta^{(m)}}^2 \boldsymbol{I}_{D_m \times D_m})$.

Theorem 1. *We define the unconditional variance-covariance of $\boldsymbol{Y}_{.j}^{(1)}$ and $\boldsymbol{Y}_{.j}^{(2)}$ as $\sum_{\mathbf{12}}^{**}$. Then we have the correlation coefficients, $\rho_{ik}, i = 1, 2, .., D_1$ and $k = 1, 2, \ldots, D_1$, constructed from $\sum_{\mathbf{12}}^{**}$, satisfy $|\rho_{ik}| < \omega$, where $\omega < 1$. In addition, the canonical correlation coefficients $|\varphi_i| < \psi$, where $\psi < 1$. The detailed proof is in Appendix A.*

Corollary 3.1. Correlation coefficients and canonical correlation coefficients calculated from the raw count data $\boldsymbol{Y}_{.j}^{(1)}$ and $\boldsymbol{Y}_{.j}^{(2)}$ will be smaller numerically in magnitude than the correlation coefficients and canonical correlation coefficients calculated from the natural parameters $\boldsymbol{\theta}_{.j}^{(1)}$ and $\boldsymbol{\theta}_{.j}^{(2)}$.

4 Simulation

4.1 Settings

In this section we conduct simulations to assess the performance of PSCCA in comparison with several existing methods that have been proposed for probabilistic correlation analysis (PCAN) (Zoh et al., 2016) , Bayesian CCA (BCCA) (Klami et al., 2013), and sparse CCA (PMD) (Witten and Tibshirani, 2009), as mentioned in the Introduction. First, we evaluate the performance of PSCCA on correlation analysis, comparing with PCAN because the PCAN paper already demonstrated that it outperforms traditional Spearman and Pearson correlation methods. Second, we compare PSCCA's performance on the canonical correlation analysis with that of BCCA, PMD, and we modified the method of PCAN based on Equations (2.5) and (2.6) to render it as an alternative approach for a probabilistic canonical correlation analysis method, which we named PCAN*.

To evaluate the methods, let $\mathbf{W}^{(\mathbf{1})*} = (\mathbf{w}_{\mathbf{1}}^{(\mathbf{1})*}, \ldots, \mathbf{w}_{\mathbf{D_1}}^{(\mathbf{1})*})$, $\mathbf{W}^{(\mathbf{2})*} = (\mathbf{w}_{\mathbf{1}}^{(\mathbf{2})*}, \ldots, \mathbf{w}_{\mathbf{D_2}}^{(\mathbf{2})*})$ be the true generated loading matrices. For the all methods with estimates $\hat{\mathbf{W}}^{(\mathbf{1})} = (\hat{\mathbf{w}}_{\mathbf{1}}^{(\mathbf{1})}, \ldots, \hat{\mathbf{w}}_{\mathbf{D_1}}^{(\mathbf{1})})$, $\hat{\mathbf{W}}^{(\mathbf{2})} = (\hat{\mathbf{w}}_{\mathbf{1}}^{(\mathbf{2})}, \ldots, \hat{\mathbf{w}}_{\mathbf{D_2}}^{(\mathbf{2})})$, we can calculate the correlations and the canonical correlations according to Equations (2.5) and (2.6). Let $\mathbf{U}^{D_1 \times D_1}$ be the true matrix and $\mathbf{V}^{D_2 \times D_2}$ be the estimated matrix. Then we construct the Frobenius loss function as $\sum_{i,j}(\mathbf{U_{ij}} - \mathbf{V_{ij}})^2$, assuming $D_1 < D_2$.

Scenario I: Correlation analysis . In the first scenario, we simulated 100 datasets assuming for each dataset $D_1 = 10$, $D_2 = 30$, and $N = 50$ subjects. The weight matrices are $\mathbf{W}_{D_1 \times d}^{(1)}$ and $\mathbf{W}_{D_2 \times d}^{(2)}$. We consider three correlation matrices for the natural parameters:

(a) the identity correlation matrix assuming the true $d = 0$ for $\mathbf{W}_{D_1 \times d}^{(1)}$ and $\mathbf{W}_{D_2 \times d}^{(2)}$.

(b) a correlation matrix obtained assuming $d = 5$ for $\mathbf{W}_{D_1 \times d}^{(1)}$ and $\mathbf{W}_{D_2 \times d}^{(2)}$.

(c) a correlation matrix obtained assuming $d = 10$ for $\mathbf{W}_{D_1 \times d}^{(1)}$ and $\mathbf{W}_{D_2 \times d}^{(2)}$.

We fit the PSCCA and PCAN to each of these 100 datasets assuming different dimensions of d to compute the posterior mean correlation matrices.

Scenario II: Canonical correlation analysis. In the second scenario, we simulated 100 datasets, for each dataset we set $N = 100$, $d = 10$ under low, moderate and high dimensions of $D^{(m)}$. We use three models for the correlation matrices of the the natural parameters $\boldsymbol{\theta}^{(m)}$.
Model I (Independent covariances): there is no covariance structure within each of the natural parameters $\boldsymbol{\theta}^{(m)}$.
Model II (Identity covariances): $\sum_{\boldsymbol{\theta}^{(1)}\theta^{(1)}} = \mathbf{I}$, $\sum_{\boldsymbol{\theta}^{(2)}\boldsymbol{\theta}^{(2)}} = \mathbf{I}$

Table 1: Summary of the Frobenius loss when estimating the true correlation structure for the natural parameters from the PCAN and our PSCCA model. Here, d is the value of d assumed for the true correlation matrix; d^* represents the value d assumed when fitting the model. Frobenius losses are calculated between the true correlation matrix at the natural parameter level vs the posterior mean correlation estimated based on the posterior of $\mathbf{W}^{(1)}$, $\mathbf{W}^{(2)}$ and the other parameters.

		PSCCA		PCAN	
d	d^*	Mean	95%CI	Mean	95%CI
0	2	25.21	(24.35, 26.54)	28.57	(28.31, 29.20)
0	5	19.55	(19.20, 20.49)	23.30	(21.85, 24.95)
0	10	14.12	(13.23, 14.83)	21.26	(19.66, 22.17)
5	2	4.25	(4.05, 4.34)	22.51	(21.26, 22.87)
5	5	3.89	(3.60, 4.24)	19.88	(18.72, 20.63)
5	10	4.54	(4.21, 4.91)	20.35	(19.45, 22.44)
10	2	10.31	(9.41, 12.21)	15.55	(14.68, 16.31)
10	5	8.81	(8.26, 9.35)	14.62	(13.34, 15.21)
10	10	7.85	(7.54, 8.06)	13.27	(12.82, 13.51)

Model III (Moderate covariances): $\sum_{\boldsymbol{\theta}^{(1)}\boldsymbol{\theta}^{(1)}}$ = **0.5**, $\sum_{\boldsymbol{\theta}^{(2)}\boldsymbol{\theta}^{(2)}}$ = **0.5**

(1) Low dimensions: $D_1 = D_2$ = 60, 100, 300

(2) Moderate dimensions: $D_1 = D_2$ = 500, 1000, 2000

(3) High dimensions: $D_1 = D_2$ = 3000, 4000, 5000

Model I is used in PCAN (Zoh et al., 2016), and similar models of Model II and III have been used to generate the raw data in Sparse CCA (Gao et al., 2017). We fit PSCCA, PCAN*, PMD, and BCCA to the datasets simulated under the above scenario and compute the canonical correlation matrices for the purpose of comparison.

4.2 Results

We report the Frobenius loss in Table 1 and Stein loss (Appendix Table A1) for each of the estimated correlation matrices. Stein loss is defined as $\mathbf{diag}(\mathbf{V^{-1}U}) - \mathbf{det}(\mathbf{V^{-1}U}) - D_1$ for estimating the $D_1 \times D_1$ matrix $\mathbf{V}$ and the $D_1 \times D_1$ matrix $\mathbf{U}$.

Overall, under each of the scenarios for correlation analysis, PSCCA yields a smaller Frobenius loss and a smaller Stein loss compared to PCAN, and both methods yield much smaller Frobenius losses compared to Stein losses. We found that under the true d=5, 10, when the assumed d^* is closer to the truth, PSCCA and PCAN result in smaller Frobenius losses, and smaller losses are preferred. However, when $d = 0$ we observe the opposite situation in that the closest value to the truth when d^* = 2 yields the largest Frobenius loss for PSCCA and PCAN.

We also estimate the correlations using other standard correlation estimation methods. Because we assumed $N > D_1 + D_2$, other standard correlation estimate methods are valid. We report the summary of the Frobenius loss incurred with estimating the true correlation matrices using Pearson and Spearman approaches based on the raw data in Appendix Table A2. PSCCA resulted in a smaller Frobenius loss, whereas PCAN and Spearman correlations perform similarly under the true d = 0, 10, which is consistent with the results in PCAN paper (Zoh et al., 2016).

For the canonical correlation analysis comparison, we report the summary of the results compared with PCAN*, BCCA, and PMD under Models I-III and moderate and high dimensions of D_m in Table 2.

Low dimensions. The estimator of PSCCA is closer to the truth than the estimates given by PCAN*, PMD, and BCCA. It also is worth noting that under Model II, PMD performs similarly with PSCCA and when $D_1 = D_2$ = 60, PMD performs slightly better than PSCCA because we generated the natural parameters with identity variance matrices. However, under Model I and Model III, PMD performs poorly. In other scenarios, PSCCA uniformly outperforms the three competitors. This confirms that methods with no assumptions on the variance matrices have broader applicability. Another point worth noting is that BCCA produces the largest standard errors compared to other methods, which indicates very unstable estimation for count data.

Moderate and High dimensions. PSCCA continues to outperform the competitors when the dimensions are moderate and high. PMD still displays similar performance with PSCCA under Model II. This suggests when the identity variance assumption holds, the performance of PMD can be improved. When the dimension exceeds 1000 under Model II, however, PMD displays very large standard errors. Under Model III, PCAN* performs nearly as well as PSCCA, which indicates there exists a moderate level of correlation and the canonical correlation estimated from the natural parameters is closer to the truth. BCCA still has large standard errors compared to other methods, which indicates it is not a proper method for sparse count data.

Table 2: Simulation results for Models I-III under low, moderate and high dimensions of D_m. The reported numbers are the medians and standard errors (in parentheses) of canonical correlation's Frobenius loss over 100 replicates. Best performer is highlighted in bold.

	$D_1 = D_2 = 60$			$D_1 = D_2 = 100$			$D_1 = D_2 = 300$		
	Mode I	Mode II	Model III	Model I	Model II	Model III	Model I	Model II	Model III
PSCCA	**0.8178**	1.102	**0.8706**	**1.142**	**1.287**	**1.082**	**1.263**	**1.124**	**1.349**
	(0.1321)	(0.075)	(0.0191)	(0.037)	(0.1083)	(0.0652)	(0.014)	(0.0152)	(0.0352)
PCAN*	1.258	1.115	1.135	1.597	1.347	1.272	1.594	1.295	1.726
	(0.1459)	(0.0649)	(0.0102)	(0.1710)	(0.028)	(0.0687)	(0.022)	(0.0159)	(0.0913)
PMD	1.246	**1.094**	1.569	1.445	1.293	1.450	1.495	1.135	1.742
	(0.049)	(0.0344)	(0.0342)	(0.033)	(0.0253)	(0.4562)	(0.009)	(0.0235)	(0.4289)
BCCA	1.436	1.181	1.035	1.328	1.362	1.156	1.292	1.7997	1.543
	(0.2462)	(0.2739)	(0.2268)	(0.3551)	(0.281)	(0.1596)	(0.276)	(0.0197)	(0.2039)
	$D_1 = D_2 = 500$			$D_1 = D_2 = 1000$			$D_1 = D_2 = 2000$		
	Mode I	Mode II	Model III	Model I	Model II	Model III	Model I	Model II	Model III
PSCCA	**1.124**	**1.307**	**1.381**	**1.448**	**1.265**	**1.611**	**1.220**	**1.532**	**1.542**
	(0.0076)	(0.0570)	(0.0158)	(0.0014)	(0.0104)	(0.058)	(0.0020)	(0.0368)	(0.0267)
PCAN*	1.737	1.407	1.547	1.606	1.842	1.661	1.609	1.842	1.589
	(0.0138)	(0.0323)	(0.0152)	(0.0043)	(0.0302)	(0.0934)	(0.0017)	(0.0302)	(0.0640)
PMD	1.471	1.375	1.681	1.518	1.684	1.694	1.263	1.591	1.783
	(0.0037)	(0.0121)	(0.5264)	(0.0043)	(0.5903)	(0.4157)	(0.0050)	(0.4013)	(0.4106)
BCCA	1.332	1.497	1.589	1.603	1.717	1.788	1.391	1.669	1.969
	(0.4000)	(0.4430)	(0.2452)	(0.3021)	(0.1845)	(0.2262)	(0.3900)	(0.2331)	(0.3903)
	$D_1 = D_2 = 3000$			$D_1 = D_2 = 4000$			$D_1 = D_2 = 5000$		
	Mode I	Mode II	Model III	Model I	Model II	Model III	Model I	Model II	Model III
PSCCA	**1.031**	**1.120**	**1.210**	**1.231**	**1.425**	1.463	**1.340**	1.437	**1.400**
	(0.0431)	(0.0262)	(0.0323)	(0.0352)	(0.0162)	(0.0320)	(0.0252)	(0.0370)	(0.0280)
PCAN*	1.627	1.580	1.343	1.581	1.723	1.480	1.710	1.783	1.426
	(0.0422)	(0.0212)	(0.0320)	(0.0433)	(0.0202)	(0.0624)	(0.0176)	(0.0252)	(0.0432)
PMD	1.586	1.224	1.315	1.453	1.479	**1.436**	1.556	**1.410**	1.487
	(0.0265)	(0.4280)	(0.0372)	(0.0363)	(0.5032)	(0.0176)	(0.0387)	(0.7032)	(0.0642)
BCCA	1.420	1.454	1.503	1.530	1.566	1.688	1.460	1.568	1.730
	(0.2360)	(0.5120)	(0.3412)	(0.2761)	(0.3142)	(0.2786)	(0.4320)	(0.3650)	(0.4532)

5 Real Data Analysis

The Cancer Genome Atlas (TCGA) (http://cancergenome.nih.gov) was initiated in 2006 to develop a publicly accessible infrastructure data on an increasing number of well-characterized cancer genomes. TCGA finalized tissue collection with matched tumor and normal tissues from 11,000 patients with 33 cancer types and subtypes, including 10 rare types of cancer. TCGA data has been used to characterize key genomic changes, find novel mutations, define intrinsic tumor types, discover similarities and differences across cancer types, reveal therapy resistance mechanisms, and collect tumor evolution evidence (Tomczak et al., 2015).

MicroRNAs (miRNA) are very short non-coding RNAs that regulate gene expression at the post-transcriptional level. They bind to messenger RNA (mRNA) and inhibit translation or induce mRNA degradation. There are many studies that demonstrate negative correlations in the expression of specific miRNA and their corresponding target mRNA, and their interaction in many disease-related regulatory pathways is well established (Ruike et al., 2008; Wang and Li, 2009; Shah et al., 2011). In recent years, there have been numerous studies about miRNA and mRNA correlation analysis on different cancer diseases using TCGA data (Ding et al., 2019; Yu et al., 2019). However, these correlation analyses all are based on the normalized continuous data, not the raw count data.

In our analysis, we consider the read count NGS expression data from squamous cell lung cancer (LUSC). We downloaded the LUSC dataset from TCGA data portal. LUSC has 504 samples, and we processed the tumor miRNA and mRNA data according to TCGAbiolinks (Colaprico et al., 2016). Each sample contains 1,881 miRNA and 56,537 mRNA. Firstly, we are interested in the correlation analysis between low-expressed mRNA and a given set of miRNA. We consider $N = 100$ matched miRNA and mRNA samples, and we select $D_1 = 50$ miRNA and $D_2 = 60$ mRNA. For miRNA we choose some reported with high correlations with mRNA in PCAN (Zoh et al., 2016), and among 60 mRNA we choose 30 mRNA with average counts between 1 and 2, and the remaining 30 mRNA are the significant expressed genes reported by (Shah et al., 2011). We fit the PSCCA model using the priors in Appendix B under $d = 2, 5, 10$. We ran two separate MCMC chains for 10,000 iterations, and monitored them for proper mixing. The first 5,000 iterations were discarded as burn-in and the inference was based on the 10,000 remaining iterations. We estimate the correlations between miRNA and mRNA based on the posterior mean values of natural parameters, and we also report results based on the standard correlation estimation approaches (Spearman and Pearson) applied to the raw data for comparison. We display, as a heatmap, the posterior mean estimates of each correlation from PSCCA in Figure 1, and PCAN in Appendix Figure A1.

The correlation results identify very interesting miRNA-mRNA interactions. PSCCA demonstrated the highest power to select the potentially correct miRNA-mRNA interactions. Our results show that miR-539 is negatively correlated with genes $RPS26P49$, $KRT18P37$, $TP63$, and $CA1$. The gene encoding miR-539 is located on human chromosome 4q32.31, and miR-539 has been reported to be down-regulated in many human cancers, including prostate cancer, nasopharyngeal carcinoma and thyroid cancer, and miR-539 has been reported to play a tumor suppression role in many human malignancies (Guo and Gong, 2018). Through TargetScanHuman we confirmed that $TP63$ is the target gene of miR-539, and all the methods estimated the correct negative correlation direction. However, PSCCA has the lowest estimated correlation value -0.3102 between miR-539

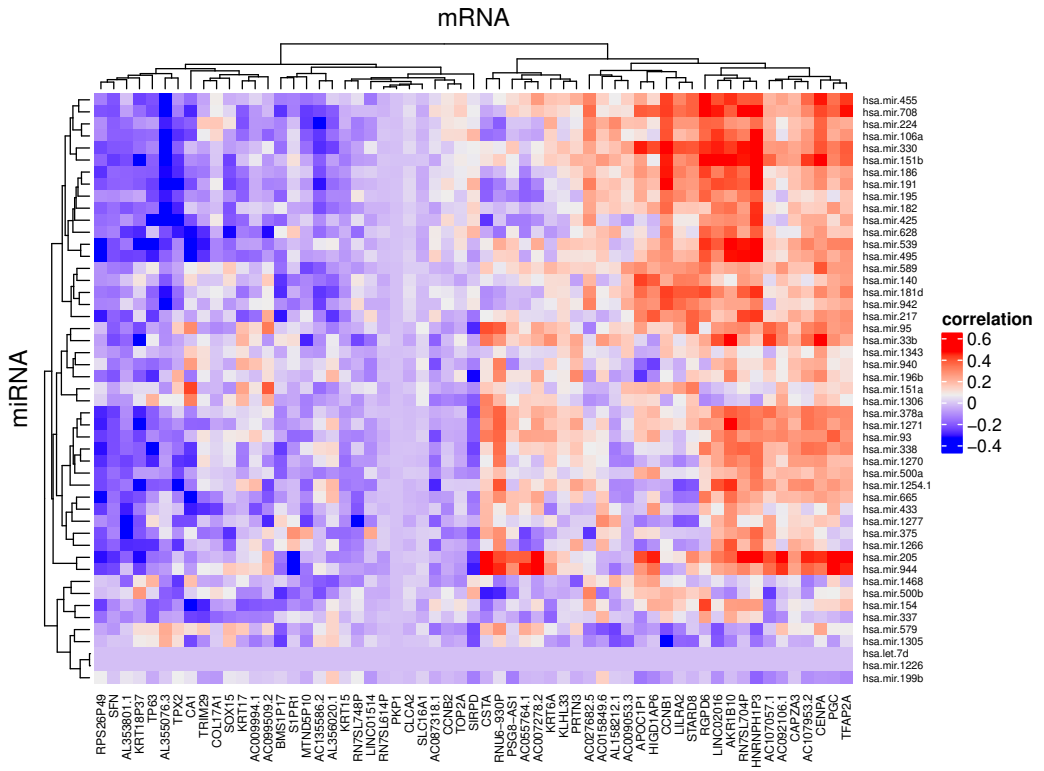


Figure 1: *PSCCA heatmap of the posterior mean correlation estimates between the miRNA and mRNA under* $d = 10$. *Red color indicates the positive correlation and blue color indicates the negative correlation.*

and $TP63$ compared to -0.1054 in PCAN, -0.1149 in Spearman, and -0.1218 in Pearson. Meanwhile, from TargetScanHuman, we noticed that miR-539 also regulates $KRT13$, $CA11$ which are the same protein family of $KRT18P37$, and $CA1$, respectively. Thus, our findings might add new members for the target gene family of miR-539, and provide more clues for miR-539's regulation role in lung cancer. Another interesting miRNA is miR-205 which was reported to play a dual role, depending on the specific tumor type and target genes (Nordby et al., 2017), we found it to be negatively correlated with $S1PR1$, $RPS26P49$, SFN, and $SLC16A1$ in our study. $S1PR1$ is the target validated through TargetScanHuman, and the estimated correlation value between $S1PR1$ and miR-205 in PSCCA is -0.2875 compared to -0.0896 in PCAN, -0.00617 in Spearman, and -0.01743 in Pearson. The same situation occurred for the mir-338 and TP63 interaction in which all the methods estimated the correct negative correlation directions, but PSCCA has the most extreme negative value compared to other three methods. Here, we report a few interesting miRNA and their

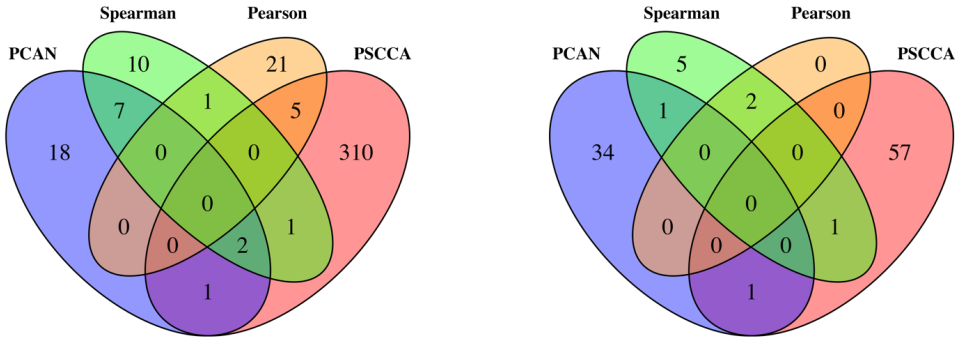
estimated correlation with mRNA in Table 3. From Table 3, we can see that for the same pair of miRNA-mRNA that our PSCCA can estimate the most extreme correlation values among all the other methods.

Table 3: Correlation estimates on LUSC data from PSCCA, PCAN, Spearman, and Pearson under $d = 10$.

miRNA	mRNA	PSCCA	PCAN	Spearman	Pearson
hsa-mir-205	RPS26P49	-0.2690	-0.1677	-0.1335	-0.1037
hsa-mir-205	S1PR1	-0.2875	-0.0896	-0.0062	-0.0174
hsa-mir-205	KRT18P37	-0.2742	0.0471	0.0759	-0.0173
hsa-mir-1305	CCBN1	-0.2852	0.0634	-0.1068	-0.0814
hsa-mir-375	AL353801.1	-0.3316	-0.1101	-0.1553	-0.0299
hsa-mir-330	APOC1P1	0.4297	0.0050	0.0271	0.0456
hsa-mir-338	TP63	-0.3017	-0.0137	-0.0385	-0.0166
hsa-mir-338	RGPD6	0.3062	0.1149	0.1469	0.1893
hsa-mir-539	TP63	-0.3102	-0.1054	-0.1149	-0.1218
hsa-mir-539	RSP26P49	-0.2949	-0.0769	-0.0527	-0.0427
hsa-mir-539	CA1	-0.3046	-0.1226	-0.1629	-0.1058
hsa-mir-338	CA1	-0.2267	0.1559	0.1014	-0.0031

Figure 2 (a) and Figure 2 (b) show the Venn diagram depicting the overlap between the pairs miRNA-mRNA correlation idenfied by PSCCA, PCAN, Spearman and Pearson approaches. Overall, there are very few pairs jointly identified by four approaches, PSCCA identified 59 negative correlation pairs with estimated correlation values less than -0.25, compared to 36 in PCAN, 9 in Spearman, and 2 in Pearson. For positive correlations, PSCCA identified 319 pairs with estimated correlation values larger than 0.25, compared to 28 in PCAN, 27 in Pearson, and 21 in Spearman. These results suggest PSCCA can identify more extreme negative and positive correlation estimates in sparse count data, and thus have less chance to miss the true correlation pairs for high-dimensional exploratory analysis. For detailed results please check the Appendix tables A3, A4, A5.

For canonical correlation analysis, we apply PSCCA, PCAN*, PMD, and BCCA on the same LUSC dataset as above for sample size $N = 100$, $D_1 = 50$ miRNA, and $D_2 = 60$ mRNA. Here, for ease of presentation, we focus only on the first two canonical correlations. We fit all the models on $d = 2, 5, 10$, that is, the number of canonical vectors to be obtained. For PMD, we use equal tuning parameters $\lambda_{a_1} = \lambda_{\beta_1}$, in which the tuning parameter is chosen by the function CCA.permute in the R package PMA. For BCCA, we use the default settings for initial parameter values. Here, we display the results on the low and moderate dimension dataset in Table 4. PSCCA and PCAN* yield high canonical correlations, while PMD and BCCA do not perform very well, both yielding small canonical correlations. In addition, under $d = 2$ PCAN* has slightly larger canonical correlation



(a) Positive correlation

(b) Negative correlation

Figure 2: *Venn diagram summarising the positive correlation greater than 0.25 (a) and negative correlation less than -0.25 (b) under PSCCA, PCAN, Spearman, and Pearson.*

values than PSCCA, which might be because the small d cannot capture the variance about the truth. To check the reasons for poor performance of PMD and BCCA, we found that the variance matrices of the two data sets $\mathbf{X}^{(1)}$ and $\mathbf{Y}^{(2)}$ are distant from identity matrices, which severely violates the identity variance assumption imposed by PMD. Also, the data are very sparse because we selected 30 out of 60 mRNAs with the average count between 1 and 2 and the remaining 30 mRNAs have large count values which severely violates the standard normal distribution assumption on the two data sets made by BCCA. PMD and BCCA still can run on this low-dimensional dataset; however, when we apply those methods on the miRNA and mRNA data sets with moderate dimensions ($D_1 = D_2 = 1000$), the R program of PMD was shut down directly, while BCCA yields the first canonical correlation value of 0.4174 which is far less than PSCCA 0.9534 under $d = 10$. That again indicates that for high-dimensional sparse count data in genomics, efficient and more accurate canonical correlation methods are needed.

6 Discussion and Future Work

We proposed a probabilistic approach of correlation and canonical correlation analysis for sparse count data. PSCCA is a model-based approach to estimate correlations and canonical correlations at the natural parameter level rather than at the raw data level. Both the simulation study results and the real data application indicate that PSCCA compares favorably to existing methods.

We provided a theoretical justification to prove that correlation coefficients and canonical correlation coefficients calculated from the raw count data $\mathbf{X}^{(1)}$ and $\mathbf{Y}^{(2)}$ will be smaller in magnitude than the correlation coefficients and canonical correlation coefficients calculated from the natural

Table 4: Canonical correlation results on real data. d is the value of d assumed for the true correlation matrix. We report the 1st and 2nd canonical correlations. Best performer is highlighted in bold.

	d=10		d=5		d=2	
$D_1 = 50, D_2 = 60$	1st	2nd	1st	2nd	1st	2nd
PSCCA	**0.9759**	**0.9229**	**0.8732**	**0.8245**	0.7979	0.6791
PCAN*	0.8884	0.8528	0.8343	0.7733	**0.8188**	**0.7179**
PMD	0.5858	0.5368	0.5858	0.5368	0.5858	0.5147
BCCA	0.2624	0.2550	0.1981	0.1786	0.1608	0.0069
	d=10		d=5		d=2	
$D_1 = 1000, D_2 = 1000$	1st	2nd	1st	2nd	1st	2nd
PSCCA	**0.9534**	**0.9319**	**0.8643**	**0.8431**	**0.8129**	**0.7351**
PCAN*	0.9213	0.8834	0.8265	0.8014	0.7834	0.7021
PMD	-	-	-	-	-	-
BCCA	0.4174	0.4054	0.3653	0.3186	0.2112	0.1406

parameters θ and λ in section 3. And this explains why PSCCA achieves more extreme correlation and canonical correlation estimations in real data application. Meanwhile, we demonstrate that horseshoe prior can handle the sparsity very well and this, probably, is due to the horseshoe prior not regularizing the parameters far from zero, which is very important in extracting the important variables that only strongly identified in the NGS data.

As the demand increases for integrative high-dimensional complex data analysis, PSCCA is a linear method which may not be appropriate for fitting the complicated nonlinear situations. Recently, researchers in computer science and engineering developed deep CCA (Wang et al., 2015) to extract the nonlinear associations and extended it to multiple views. However, deep CCA benefits from the expressive power of deep neural networks, which have a black box drawback in that they are not easy to interpret and understand. PSCCA is a model-based approach which estimates the dependency within and between two datasets as a joint task, thus, PSCCA is more interpretable by checking the weight for each feature. One direction for our work is to develop a probabilistic deep CCA to handle more complex data structures and provide interpretable results.

References

Archambeau, C. and Bach, F. (2008), "Sparse Probabilistic Projections," *Advances in Neural Information Processing Systems 21*, 73–80.

Armagan, A., Dunson, D., and Lee, J. (2013), "Generalized double Pareto shrinkage," *Statistica Sinica*, 23(1), 119–143.

Avants, B. B., Cook, P., Ungar, L., Gee, J., and Grossman, M. (2010), "Dementia induces correlated reductions in white matter integrity and cortical thickness: a multivariate neuroimaging study with sparse canonical correlation analysis," *Neuroimage*, 50(3), 1004–1016.

Bach, F. and Jordan, M. (2006), "A probability interpretation of canonical correlation analysis," Tech. rep., University of California,Berkeley, California.

Bhattacharya, A. and Dunson, D. (2011), "Sparse Bayesian infinite factor models," *Biometrika*, 98(2), 291–306.

Bhattacharya, A., Pati, D., Pillai, N., and Dunson, D. (2014), "Dirichlet Laplace priors for optimal shrinkage," *Journal of the American Statistical Association*, 110(512), 1479–1490.

Carvalho, C., Chang, J., Lucas, J., Nevins, J., Wang, Q., and West, M. (2008), "High-dimensional sparse factor modeling: Applications in gene expression genomics," *Journal of the American Statistical Association*, 103(484), 1438–1456.

Carvalho, C., Polson, N., and Scott, J. (2010), "The horseshoe estimator for sparse signals," *Biometrika*, 97(2), 465–480.

Colaprico, A., Silva, T., Olsen, C., Garofano, L., Cava, C., Garolini, D., Sabedot, T., Malta, T., Pagnotta, S., Castiglioni, I., Ceccarelli, M., Bontempi, G., and Noushmehr, H. (2016), "TCGAbiolinks: An R/Bioconductor package for integrative analysis of TCGA data," *Nucleic Acids Research*, 44(8), e71.

Datta, J. and Ghosh, J. (2013), "Asymptotic properties of bayes risk for the horseshoe prior," *Bayesian Analysis*, 8(1), 111–132.

Ding, L., Feng, Z., and Bai, Y. (2019), "Clustering analysis of microRNA and mRNA expression data from TCGA using maximum edge-weighted matching algorithm," *BMC Medical Genomics*, 12(17), 117.

Fujiwara, Y., Miyawaki, Y., and Kamitani, Y. (2009), "Estimating image bases for visual image reconstruction from human brain activity," *Advances in Neural Information Processing Systems 22*, 576–584.

Gao, C., Ma, Z., and Zhou, H. H. (2017), "Sparse cca: Adaptive estimation and computational barriers," *The Annals of Statistics*, 45(5), 2074–2101.

George, E. and McCulloch, R. (1993), "Variable Selection via Gibbs Sampling," *Journal of the American Statistical Association*, 88(423), 881–889.

Geweke, J. and Zhou, G. (1996), "Measuring the pricing error of the arbitrage pricing theory," *Review of Financial Studies*, 9(2), 557–587.

Guo, J. and Gong, G.H.and Zhang, B. (2018), "miR-539 acts as a tumor suppressor by targeting epidermal growth factor receptor in breast cancer," *Scientific Reports*, 8(1), 2073.

Hardoon, D. R. and Shawe-Taylor, J. (2011), "Sparse canonical correlation analysis," *Machine Learning*, 83(3), 331–353.

Hotelling, H. (1936), "Relations between two sets of variates," *Biometrika*, 28(3-4), 321–377.

Huopaniemi, I., Suvitaival, T., Nikkila, J., Oresic, M., and Samuel, K. (2010), "Multivariate multi-way analysis of multi-source data," *Bioinformatics*, 26(12), i391–i398.

Klami, A., Virtanen, S., and Kaski, S. (2013), "Bayesian Canonical Correlation Analysis," *Journal of Machine Learning Research*, 14, 965–1003.

Knowles, D. and Ghahramani, Z. (2011), "Nonparametric Bayesian sparse factor models with application to gene expression modeling," *The Annals of Applied Statistics*, 5(2B), 1534–1552.

Lopes, H. and West, M. (2004), "Bayesian model assessment in factor analysis," *Statistica Sinica*, 14(1), 41–68.

Lu, Y., Wang, B., Jiang, F., Mo, X., Wu, L., He, P., Lu, X., Deng, F., and Lei, S. (2019), "Multi omics integrative analysis identified SNP methylation mRNA: Interaction in peripheral blood mononuclear cells," *Journal of Cellular and Molecular Medicine*, 23(7), 4601–4610.

Muirhead, R. (1982), *Aspects of Multivariate Statistical Theory*, John Wiley.

Neal, R. (1996), *Bayesian Learning for Neural Networks*, Springer-Verlag.

Nordby, Y., Richardsen, E., Ness, N., Donnem, T., Patel, H., Busund, L., Bremnes, R., and Andersen, S. (2017), "High miR-205 expression in normal epithelium is associated with biochemical failure-an argument for epithelial crosstalk in prostate cancer?" *Scientific Reports*, 7, 16308.

Pas, S., Kleijin, B., and Vaart, A. (2014), "The horseshoe estimator: Posterior concentration around nearly black vectors," *Electronic Journal of Statistics*, 8, 2585–2618.

Pas, S., Szabo, B., and Vaart, A. (2017), "Adaptive posterior contraction rates for the horseshoe," *Electronic Journal of Statistics*, 11, 3196–3225.

Piironen, J. and Vehtari, A. (2017), "Sparsity of information and regularization in the horseshoe and other shrinkage priors," *Electronic Journal of Statistics*, 11, 5018–5051.

Polson, N. and Scott, J. (2011), "Shrink globally, act locally: Sparse Bayesian regularization and prediction," *In Bayesian Statistics 9, eds. J.M. Bernardo et al.,*, 501–538.

Ruike, Y., Ichimura, A., Tsuchiya, S., Shimizu, K., Kunimoto, R., Okuno, Y., and Tsujimoto, G. (2008), "Global correlation analysis for micro-RNA and mRNA expression profiles in human cell lines," *Journal of Human Genetics*, 53(6), 515–523.

Safo, S., Li, S., and Long, Q. (2018), "Integrative analysis of transcriptomic and metabolomic data via sparse canonical correlation analysis with incorporation of biological information," *Biometrics*, 74(1), 300–312.

Shah, M., Schwartz, S., Zhao, C., Davidson, L., Zhou, B., Lupton, J., Ivanov, I., and Chapkin, R. (2011), "Integrated microrna and mrna expression profiling in a rat colon carcinogenesis model: effect of a chemo-protective diet," *Physiological genomics*, 43(10), 640.

Tipping, M. (2001), "Sparse Bayesian Learning and the Relevance Vector Machine," *Journal of Machine Learning Research*, 1, 211–244.

Tomczak, K., Czerwińska, P., and Wiznerowicz, M. (2015), "The Cancer genome atlas (TCGA): an immeasurable source of knowledge," *Contemporary Oncology*, 19(1A), A68–77.

Virtanen, S., Klami, A., and Kaski, S. (2011), "Bayesian cca via group sparsity," *Proceedings of the 28th International Conference on International Conference on Machine Learning*, 457–464.

Wang, W., Arora, R., Livescu, K., and Bilmes, J. (2015), "On Deep Multi-View Representation Learning," *International Conference on Machine Learning*, 1083–1092.

Wang, Y. P. and Li, K. (2009), "Correlation of expression profiles between microRNAs and mRNA targets using NCI-60 data," *BMC Genomics*, 10, 218.

West, M. (2003), "Bayesian factor regression models in the "large p, small n" paradigm," *In Bayesian Statistics 7, eds. J.M. Bernardo et al.,*, 723–732.

Witten, D. M. and Tibshirani, R. (2009), "A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis," *Biostatistics*, 10(3), 515–534.

Yu, N., Yong, S., Choi, Y., Jung, Y., Kim, D., Seo, J., Lee, Y., Baek, D., Lee, J., Lee, S., Lee, J., Kim, J., Kim, J., and Lee, S. (2019), "Identification of tumor suppressor miRNAs by integrative miRNA and mRNA sequencing of matched tumor-normal samples in lung adenocarcinoma," *Molcular Oncology*, 13, 1356–1368.

Zhao, S., Gao, C., Mukherjee, S., and Engelhardt, B. E. (2016), "Bayesian group factor analysis with structured sparsity," *Journal of Machine Learning Research*, 17, 1–47.

Zoh, R., Mallick, B., Ivanov, I., Baladandayuthapani, V., Manyam, G., Chapkin, R., Lampe, J., and Carroll, R. (2016), "PCAN: Probabilistic Correlation Analysis of Two Non-normal Data Sets," *Biometrics*, 72(4), 1358–1368.

Received: *21 March 2022*

Accepted: *15 July 2022*

Appendix

A Proof of Theorem 1

Let $F(\cdot|\cdot)$ be a cumulative distribution function from the natural parameter exponential family as we discussed in the model section. We model

$$Y_{ij}^{(m)} \sim F(Y|\theta_{ij}^{(m)}) = \text{Poisson}(y|\theta_{ij}^{(m)}), \tag{A.1}$$

where $m = 1, 2$, $i = 1, 2, \ldots, D_m$, $j = 1, 2, \ldots, N$. Let $\boldsymbol{Y}_{.j}^{(m)} = [Y_{1j}^{(m)}, Y_{2j}^{(m)}, \ldots, Y_{D_m j}^{(m)}]^T$ and $\boldsymbol{\theta}_{.j}^{(m)} = [\theta_{1j}^{(m)}, \theta_{2j}^{(m)}, \ldots, \theta_{D_m j}^{(m)}]^T$. Then, as in Equations (2.1) and (2.2),

$$\boldsymbol{\theta}_{.j} = \boldsymbol{\mu}_{\theta}^{(m)} + \boldsymbol{W}^{(m)}\boldsymbol{Z}_j + \boldsymbol{\epsilon}_{.j}^{(m)}$$

where $\boldsymbol{W}^{(m)} \in \mathbb{R}^{D_m \times d}$, $\boldsymbol{Z}_j \sim N_d(\mathbf{0}_{d\times d}, \boldsymbol{I}_{d\times d})$ and $\boldsymbol{\epsilon}_{.j}^{(m)} \sim N_{D_m}(\mathbf{0}_{D_m\times D_m}, {\sigma_{\theta}^{(m)}}^2 \boldsymbol{I}_{D_m\times D_m})$.

$$\boldsymbol{\theta}_{.j} = \boldsymbol{\mu}_{\theta}^{(m)} + \boldsymbol{W}^{(m)}\boldsymbol{Z}_j + \boldsymbol{\epsilon}_{.j}^{(m)}$$

where $\boldsymbol{W}^{(m)} \in \mathbb{R}^{D_m \times d}$, $\boldsymbol{Z}_j \sim N_d(\mathbf{0}_{d\times d}, \boldsymbol{I}_{d\times d})$ and $\boldsymbol{\epsilon}_{.j}^{(m)} \sim N_{D_m}(\mathbf{0}_{D_m\times D_m}, {\sigma_{\theta}^{(m)}}^2 \boldsymbol{I}_{D_m\times D_m})$.

Then the vector of the natural parameters follows a multivariate normal distribution and its expectation vector and variance-covariance matrix are

$$\boldsymbol{\mu} = E\begin{bmatrix}\boldsymbol{\theta}^{(1)} \\ \boldsymbol{\theta}^{(2)}\end{bmatrix} = \begin{bmatrix}\boldsymbol{\mu}_{\boldsymbol{\theta}^{(1)}} \\ \boldsymbol{\mu}_{\boldsymbol{\theta}^{(2)}}\end{bmatrix}$$

$$\boldsymbol{\Sigma} = Var\begin{bmatrix}\boldsymbol{\theta}^{(1)} \\ \boldsymbol{\theta}^{(2)}\end{bmatrix} = \begin{pmatrix} \boldsymbol{W}^1\boldsymbol{W}^{\boldsymbol{1}T} + {\sigma_{\theta}^{(1)}}^2\boldsymbol{I}_{D_1\times D_1} & \boldsymbol{W}^1\boldsymbol{W}^{\boldsymbol{2}T} \\ \boldsymbol{W}^2\boldsymbol{W}^{\boldsymbol{1}T} & \boldsymbol{W}^2\boldsymbol{W}^{\boldsymbol{2}T} + {\sigma_{\theta}^{(2)}}^2\boldsymbol{I}_{D_2\times D_2}\end{pmatrix}$$

In turn, the exponentiated natural parameters follow a multivariate lognormal distribution. Using the results described in Skoulakis (2008), we provide expressions for the expectation vector and variance-covariance matrix of the exponentiated natural parameters. We need additional notation, however, for this purpose. Let

(a) vecdiag $(\mathbf{C})$ denotes the $t \times 1$ vector of the diagonal elements of the $t \times t$ matrix $\mathbf{C}$

(b) $\mathbf{1_{b\times b}}$ denote and $a \times b$ matrix of unit values

(c) $\odot$ denote elementwise multiplication of two matrices with the same dimensions

Then

$$\boldsymbol{\mu^*} = \begin{bmatrix}\boldsymbol{\mu}^*_{\theta^{(1)}} \\ \boldsymbol{\mu}^*_{\theta^{(2)}}\end{bmatrix} = E\begin{bmatrix}exp(\boldsymbol{\theta^{(1)}}) \\ exp(\boldsymbol{\theta^{(2)}})\end{bmatrix} = \begin{bmatrix}exp\left\{\boldsymbol{\mu}_{\boldsymbol{\theta}^{(1)}} + \frac{1}{2}\,\text{vecdiag}\left(\boldsymbol{W}^1\boldsymbol{W}^{\boldsymbol{1}T} + {\sigma_{\theta}^{(1)}}^2\boldsymbol{I}_{D_1\times D_1}\right)\right\} \\ exp\left\{\boldsymbol{\mu}_{\boldsymbol{\theta}^{(2)}} + \frac{1}{2}\,\text{vecdiag}\left(\boldsymbol{W}^2\boldsymbol{W}^{\boldsymbol{2}T} + {\sigma_{\theta}^{(2)}}^2\boldsymbol{I}_{D_2\times D_2}\right)\right\}\end{bmatrix}$$

$$\Sigma = \begin{pmatrix} \boldsymbol{\Sigma}^*_{\theta^{(1)}\theta^{(1)}} & \boldsymbol{\Sigma}^*_{\theta^{(1)}\theta^{(2)}} \\ \boldsymbol{\Sigma}^*_{\theta^{(2)}\theta^{(1)}} & \boldsymbol{\Sigma}^*_{\theta^{(2)}\theta^{(2)}} \end{pmatrix} = Var \begin{bmatrix} exp(\boldsymbol{\theta^{(1)}}) \\ exp(\boldsymbol{\theta^{(2)}}) \end{bmatrix} = (\boldsymbol{\mu^* \mu^{*T}}) \odot \left\{ exp(\boldsymbol{\Sigma}) - \mathbf{1}_{(D_1+D_2)\times(D_1+D_2)} \right\}$$

We now are in position to determine the unconditional expectation vector and unconditional variance-covariance matrix of the multivariate Poisson distribution of $\boldsymbol{Y}^{(1)}$ and $\boldsymbol{Y}^{(2)}$. The expectation vector is

$$\boldsymbol{\mu^{**}} = \begin{bmatrix} \boldsymbol{\mu^{**}_{\theta^{(1)}}} \\ \boldsymbol{\mu^{**}_{\theta^{(2)}}} \end{bmatrix} = E \begin{bmatrix} \boldsymbol{Y}^{(1)} \\ \boldsymbol{Y}^{(2)} \end{bmatrix} = E \left\{ E \begin{bmatrix} \boldsymbol{Y}^{(1)} | \boldsymbol{\theta}^{(1)} \\ \boldsymbol{Y}^{(2)} | \boldsymbol{\theta}^{(2)} \end{bmatrix} \right\} = E \begin{bmatrix} exp(\boldsymbol{\theta}^{(1)}) \\ exp(\boldsymbol{\theta}^{(2)}) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}^*_{\theta^{(1)}} \\ \boldsymbol{\mu}^*_{\theta^{(2)}} \end{bmatrix}$$

The variance-covariance matrix is

$$\begin{aligned} \Sigma^{**} &= \begin{pmatrix} \boldsymbol{\Sigma}^{**}_{\theta^{(1)}\theta^{(1)}} & \boldsymbol{\Sigma}^{**}_{\theta^{(1)}\theta^{(2)}} \\ \boldsymbol{\Sigma}^{**}_{\theta^{(2)}\theta^{(1)}} & \boldsymbol{\Sigma}^{**}_{\theta^{(2)}\theta^{(2)}} \end{pmatrix} = Var \begin{bmatrix} \boldsymbol{Y}^{(1)} \\ \boldsymbol{Y}^{(2)} \end{bmatrix} \\ &= Var \left\{ E \begin{bmatrix} \boldsymbol{Y}^{(1)} | \boldsymbol{\theta}^{(1)} \\ \boldsymbol{Y}^{(2)} | \boldsymbol{\theta}^{(2)} \end{bmatrix} \right\} + E \left\{ Var \begin{bmatrix} \boldsymbol{Y}^{(1)} | \boldsymbol{\theta}^{(1)} \\ \boldsymbol{Y}^{(2)} | \boldsymbol{\theta}^{(2)} \end{bmatrix} \right\} \\ &= Var \begin{bmatrix} exp(\boldsymbol{\theta}^{(1)}) \\ exp(\boldsymbol{\theta}^{(2)}) \end{bmatrix} + E \begin{pmatrix} Diag \left\{ exp(\boldsymbol{\theta}^{(1)}) \right\} & \mathbf{0}_{D_1 \times D_1} \\ \mathbf{0}_{D_2 \times D_2} & Diag \left\{ exp(\boldsymbol{\theta}^{(2)}) \right\} \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{\Sigma}^*_{\theta^{(1)}\theta^{(1)}} & \boldsymbol{\Sigma}^*_{\theta^{(1)}\theta^{(2)}} \\ \boldsymbol{\Sigma}^*_{\theta^{(2)}\theta^{(1)}} & \boldsymbol{\Sigma}^*_{\theta^{(2)}\theta^{(2)}} \end{pmatrix} + \begin{pmatrix} Diag \left\{ \boldsymbol{\mu}^*_{\boldsymbol{\theta^{(1)}}} \right\} & \mathbf{0}_{D_1 \times D_1} \\ \mathbf{0}_{D_2 \times D_2} & Diag \left\{ \boldsymbol{\mu}^*_{\boldsymbol{\theta^{(2)}}} \right\} \end{pmatrix} \end{aligned}$$

where $Diag()$ denotes the formation of a diagonal matrix from the vector argument and $\mathbf{0}_{a\times b}$ denotes an $a \times b$ matrix of zero values.

Thus, $\boldsymbol{\Sigma}^{**}$, the unconditional variance-covariance of $\boldsymbol{Y}^{(1)}$ and $\boldsymbol{Y}^{(2)}$, consists of the sum of a positive-definite matrix and a diagonal matrix. Although $\boldsymbol{\Sigma}^{**}$ itself is a positive-definite matrix, the addition of the diagonal matrix inflates the diagonal elements of $\boldsymbol{\Sigma}^{**}$. This inflation results in an off-diagonal element of $\boldsymbol{\Sigma}^{**}$, say σ^{**}_{lm}, a covariance, not capable of reaching $\sqrt{\sigma^{**}_{ll}\sigma^{**}_{mm}}$. In other words, the (l, m) correlation coefficient constructed from $\boldsymbol{\Sigma}^{**}$ will be bounded within $[-b_{lm}, b_{lm}]$, where $b_{lm} < 1$. Likewise, canonical correlation coefficients constructed from$\boldsymbol{\Sigma}^{**}$will be bounded above by a number less than 1. Hence, correlation coefficients and canonical correlation coefficients calculated from the raw count data $\boldsymbol{Y}^{(1)}$ and $\boldsymbol{Y}^{(2)}$. will be smaller numerically in magnitude than the correlation coefficients and canonical correlation coefficients calculated from the natural parameters $\boldsymbol{\theta}^{(1)}$ and $\boldsymbol{\theta}^{(2)}$.

B Priors Used for Simulation Studies

The priors used for simulation studies:

$$\begin{aligned} W_{ij}^{(1)} \sim \text{Normal}(0, 1.5) \quad if \quad i < j; W_{ij}^{(1)} \sim \text{Normal}(0, 1.5)\mathbf{1}(W_{jj}^{(1)} > 0) \quad if \quad i = j; \\ W_{kj}^{(2)} \sim \text{Normal}(0, 1.5) \quad if \quad k < j; W_{jj}^{(2)} \sim \text{Normal}(0, 1.5)\mathbf{1}(W_{jj}^{(2)} > 0) \quad if \quad k = j; \\ \mu_\theta^{(1)} = \mu_\theta^{(2)} = 1, \\ \sigma_\theta^{(1)} = \sigma_\theta^{(2)} = 0.1; \end{aligned} \tag{B.1}$$

The priors used for real data analysis PCAN:

$$\begin{aligned} W_{ij}^{(1)} \sim \text{Normal}(0, \boldsymbol{\tau}_j^{-1}) \quad if \quad i < j; W_{ij}^{(1)} \sim \text{Normal}(0, \boldsymbol{\tau}_j^{-1})\mathbf{1}(W_{jj}^{(1)} > 0) \quad if \quad i = j; \\ W_{kj}^{(2)} \sim \text{Normal}(0, \boldsymbol{\tau}_j^{-1}) \quad if \quad k < j; W_{jj}^{(2)} \sim \text{Normal}(0, \boldsymbol{\tau}_j^{-1})\mathbf{1}(W_{jj}^{(2)} > 0) \quad if \quad k = j; \\ \boldsymbol{\tau}_j \sim \text{Gamma}(2, 1); \\ \boldsymbol{\mu}_\theta^{(1)} \sim \Pi_{k=1}^{D_1}\text{Normal}(0, 5), \\ \boldsymbol{\mu}_\theta^{(2)} \sim \Pi_{i=1}^{D_2}\text{Normal}(0, 5), \\ {\sigma_\theta^{(1)}}^2 \sim \text{Inv-}\chi^2(10, 0.05), \\ {\sigma_\theta^{(2)}}^2 \sim \text{Inv-}\chi^2(10, 0.05); \end{aligned} \tag{B.2}$$

The priors used for real data analysis PSCCA:

$$\begin{aligned} W_{ij}^{(m)} \sim \text{Normal}(W_{ij}^{(1)}|0, {\lambda_i^{(1)}}^2{\tau^{(m)}}^2) \quad if \quad i < j, \\ W_{kj}^{(m)} \sim \text{Normal}(W_{jj}^{(1)}|0, {\lambda_i^{(m)}}^2{\tau^{(m)}}^2)\mathbf{1}(W_{jj}^{(m)} > 0) \quad if \quad k = j, \\ \tau^{(m)} \sim C^+(0, 1), \\ \lambda_i^{(m)} \sim C^+(0, 1), \\ \boldsymbol{\mu}_\theta^{(1)} \sim \Pi_{k=1}^{D_1}\text{Normal}(0, 5), \\ \boldsymbol{\mu}_\theta^{(2)} \sim \Pi_{i=1}^{D_2}\text{Normal}(0, 5), \\ {\sigma_\theta^{(1)}}^2 \sim \text{Inv-}\chi^2(10, 0.05), \\ {\sigma_\theta^{(2)}}^2 \sim \text{Inv-}\chi^2(10, 0.05); \end{aligned} \tag{B.3}$$

C Simulation Results and Real Data Analysis.

Table A1: *Summary of the Stein loss when estimating the true correlation structure for the natural parameters from the SPCCA and our PCAN model. Here, d is the number of latent variables assumed for the true correlation matrix; d^* represents the value of d assumed when fitting the model. Stein losses are calculated between the true correlation matrix at the natural parameter level vs the posterior mean correlation, estimated based on the posterior of $\boldsymbol{W}^{(1)}$, $\boldsymbol{W}^{(2)}$ and the other parameters. The reported numbers are the medians and 95% CI (in parentheses).*

		PCAN		PSCCA	
d	d^*	Mean	95%CI	Mean	95%CI
0	2	60.65	(58.22, 62.88)	56.62	(54.20, 58.33)
0	5	49.06	(46.68, 51.45)	45.52	(30.40, 60.63)
0	10	56.21	(31.33, 81.08)	54.27	(47.49, 61.06)
5	2	68.29	(60.75, 75.77)	14.50	(13.49, 15.51)
5	5	63.56	(57.48, 69.64)	9.098	(8.69, 9.59)
5	10	65.60	(59.35, 71.86)	11.57	(11.91, 11.98)
10	2	52.86	(49.74, 55.96)	46.19	(42.16, 50.20)
10	5	55.22	(51.05, 59.38)	37.49	(35.88, 39.11)
10	10	55.05	(50.79, 59.31)	36.56	(30.63, 42.49)

Table A2: Summary of the Frobenius losses when estimating the true correlation for the natural parameters from PSCCA, PCAN, Pearson, and Spearman. Here, d is the number of latent variables assumed for the true correlation matrix; d^* represents the value of d assumed when fitting the model. Frobenius losses are calculated between the true correlation matrix at the natural parameter level vs the posterior mean correlation, estimated based on the posterior of $\mathbf{W}^{(1)}$, $\mathbf{W}^{(2)}$ and the other parameters. The reported numbers are the medians and standard errors (in parentheses).

d	d^*	PSCCA	PCAN	Pearson	Spearman
0	10	14.12	21.26	17.89	21.50
		(1.43)	(1.57)	(0.71)	(1.23)
5	10	4.54	20.35	9.74	10.32
		(0.22)	(3.14)	(0.22)	(0.21)
10	10	7.85	13.27	11.47	12.76
		(1.08)	(3.37)	(0.42)	(0.33)

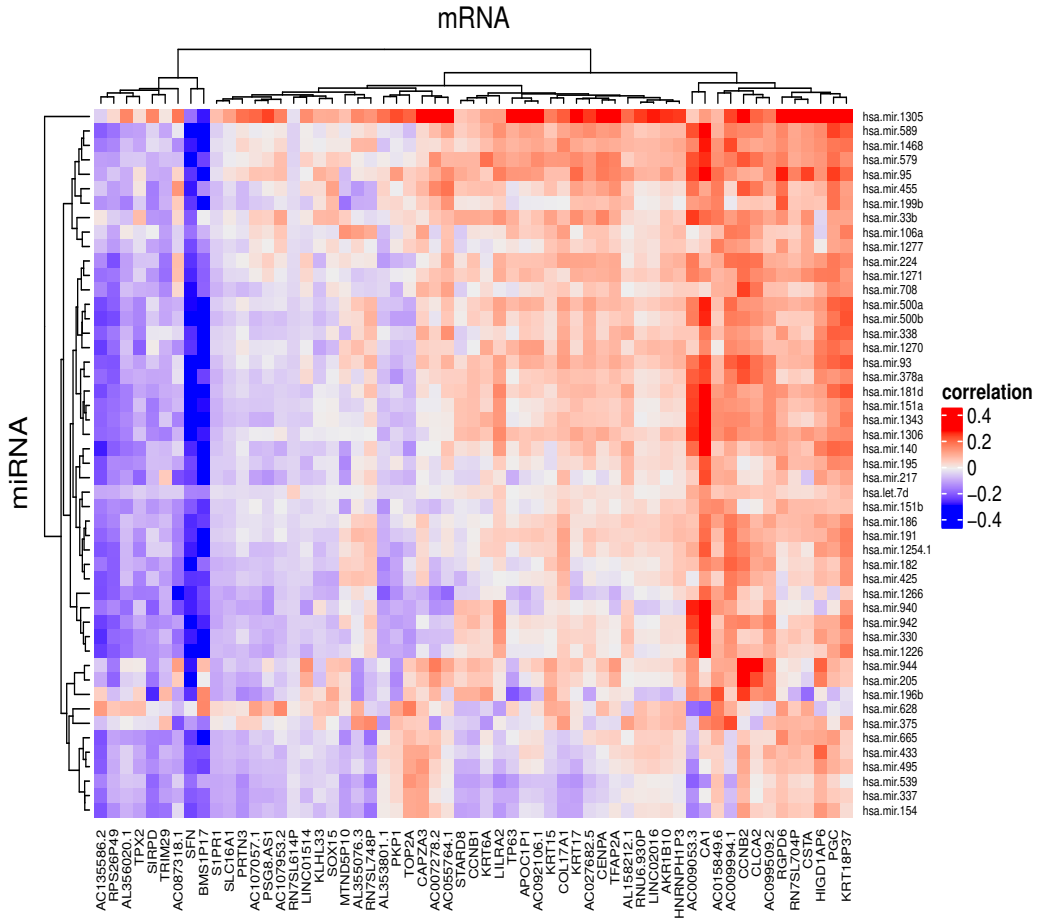


Figure A1: *PCAN heatmap of the posterior mean correlation estimates between the miRNA and mRNA under* $d = 10$. *Red color indicates the positive correlation and blue color indicates the negative correlation.*

Table A3: *Posterior mean estimates of mRNA and miRNA pairs correlation (cor) obtained from the real data analysis using PSCCA (d=10). Here we only report the top 40 pairs with negative correlations.*

mRNA	miRNA	cor	mRNA	miRNA	cor
CA1	hsa.mir.154	-0.390	AL355076.3	hsa.mir.425	-0.2905
CA1	hsa.mir.665	-0.353	RPS26P49	hsa.mir.539	-0.294
AL355076.3	hsa.mir.708	-0.338	AL355076.3	hsa.mir.182	-0.290
AL355076.3	hsa.mir.191	-0.332	KRT18P37	hsa.mir.33b	-0.288
AL353801.1	hsa.mir.375	-0.331	AL355076.3	hsa.mir.330	-0.288
AL355076.3	hsa.mir.942	-0.329	S1PR1	hsa.mir.205	-0.287
S1PR1	hsa.mir.944	-0.328	TP63	hsa.mir.425	-0.286
KRT18P37	hsa.mir.151b	-0.322	CCNB1	hsa.mir.1305	-0.286
AL355076.3	hsa.mir.455	-0.321	RPS26P49	hsa.mir.495	-0.284
KRT18P37	hsa.mir.539	-0.317	AC135586.2	hsa.mir.224	-0.284
AL355076.3	hsa.mir.151b	-0.317	AL353801.1	hsa.mir.1277	-0.284
TPX2	hsa.mir.425	-0.314	AL355076.3	hsa.mir.106a	-0.282
TPX2	hsa.mir.1254.1	-0.312	KRT18P37	hsa.mir.1254.1	-0.280
TP63	hsa.mir.539	-0.310	RPS26P49	hsa.mir.665	-0.280
TPX2	hsa.mir.191	-0.307	CA1	hsa.mir.628	-0.279
CA1	hsa.mir.539	-0.305	SIRPD	hsa.mir.196b	-0.278
KRT18P37	hsa.mir.1271	-0.302	SOX15	hsa.mir.628	-0.276
TP63	hsa.mir.338	-0.302	AL355076.3	hsa.mir.186	-0.275
RN7SL748P	hsa.mir.1277	-0.299	KRT18P37	hsa.mir.205	-0.274
CA1	hsa.mir.495	-0.295	AC099509.2	hsa.mir.628	-0.273

Table A4: *Posterior mean estimates of mRNA and miRNA pairs correlation (cor) obtained from the real data analysis using PCAN (d=10). Here we only report the pairs contain estimated correlations less than -0.25.*

mRNA	miRNA	cor	mRNA	miRNA	cor
BMS1P17	hsa.mir.140	-0.437	SFN	hsa.mir.93	-0.281
BMS1P17	hsa.mir.195	-0.393	SFN	hsa.mir.181d	-0.280
BMS1P17	hsa.mir.338	-0.384	BMS1P17	hsa.mir.191	-0.280
BMS1P17	hsa.mir.217	-0.381	SFN	hsa.mir.205	-0.276
BMS1P17	hsa.mir.181d	-0.370	SFN	hsa.mir.1306	-0.273
BMS1P17	hsa.mir.500a	-0.368	BMS1P17	hsa.mir.665	-0.271
BMS1P17	hsa.mir.1270	-0.360	SFN	hsa.mir.151a	-0.269
BMS1P17	hsa.mir.589	-0.354	SFN	hsa.mir.579	-0.267
BMS1P17	hsa.mir.330	-0.345	SFN	hsa.mir.942	-0.267
BMS1P17	hsa.mir.500b	-0.344	SFN	hsa.mir.182	-0.264
BMS1P17	hsa.mir.1226	-0.338	SFN	hsa.mir.500b	-0.261
BMS1P17	hsa.mir.1468	-0.303	SFN	hsa.mir.589	-0.260
AC087318.1	hsa.mir.1266	-0.296	SFN	hsa.mir.378a	-0.260
SFN	hsa.mir.1343	-0.291	BMS1P17	hsa.mir.95	-0.259
BMS1P17	hsa.mir.1343	-0.288	SFN	hsa.mir.1226	-0.259
BMS1P17	hsa.mir.1254.1	-0.286	SFN	hsa.mir.500a	-0.254
BMS1P17	hsa.mir.93	-0.284	SFN	hsa.mir.708	-0.253
BMS1P17	hsa.mir.199b	-0.282	BMS1P17	hsa.mir.151a	-0.251

Table A5: *Posterior mean estimates of mRNA and miRNA pairs correlation (cor) obtained from the real data analysis using Spearman (d=10) and Pearson (d=10). Here we only report the pairs contain estimated correlations less than -0.25.*

mRNA	miRNA	cor
AC135586.2	hsa.mir.500a	-0.323
RPS26P49	hsa.mir.1468	-0.286
PSG8.AS1	hsa.mir.1254.1	-0.264
AC135586.2	hsa.mir.93	-0.259
RPS26P49	hsa.mir.182	-0.258
CA1	hsa.mir.628	-0.257
CSTA	hsa.mir.196b	-0.255
BMS1P17	hsa.mir.500a	-0.251
AC135586.2	hsa.mir.425	-0.250
PSG8.AS1 (Pearson)	hsa.mir.1254.1	-0.294
AC135586.2(Pearson)	hsa.mir.500a	-0.280