

RESEARCH ARTICLE

A robust transfer learning approach for high-dimensional linear regression to support integration of multi-source gene expression data

Lulu Pan^{1‡}, Qian Gao^{2,3‡}, Kecheng Wei^{1‡}, Yongfu Yu^{1,4,5*}, Guoyou Qin^{1,4,5*}, Tong Wang^{2,3*}

1 Department of Biostatistics, School of Public Health, Fudan University, Shanghai, China, **2** Department of Health Statistics, School of Public Health, Shanxi Medical University, Taiyuan, China, **3** Key Laboratory of Coal Environmental Pathogenicity and Prevention (Shanxi Medical University), Ministry of Education, China, **4** Shanghai Institute of Infectious Disease and Biosecurity, Shanghai, China, **5** Key Laboratory of Public Health Safety of Ministry of Education, Key Laboratory for Health Technology Assessment, National Commission of Health, Fudan University, Shanghai, China

‡ These authors share first authorship on this work.

* yu@fudan.edu.cn (YY); gyqin@fudan.edu.cn (GQ); tongwang@sxmu.edu.cn (TW)



OPEN ACCESS

Citation: Pan L, Gao Q, Wei K, Yu Y, Qin G, Wang T (2025) A robust transfer learning approach for high-dimensional linear regression to support integration of multi-source gene expression data. *PLoS Comput Biol* 21(1): e1012739. <https://doi.org/10.1371/journal.pcbi.1012739>

Editor: Ilya Ioshikhes, CANADA

Received: April 28, 2024

Accepted: December 20, 2024

Published: January 10, 2025

Copyright: © 2025 Pan et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The GTEx version 8 dataset used in this work is publicly available at https://gtexportal.org/home/downloads/adult-gtex/bulk_tissue_expression. The complete list of the CNS-related genes in MODULE 137 is available at http://robotics.stanford.edu/~erans/cancer/modules/module_137. The code for the proposed method has been implemented as R functions and is available for download from [S1 Code file](#).

Funding: This work was supported by the National Natural Science Foundation of China (GQ, No 82173612; YY, No 82273730; TW, No 82073674)

Abstract

Transfer learning aims to integrate useful information from multi-source datasets to improve the learning performance of target data. This can be effectively applied in genomics when we learn the gene associations in a target tissue, and data from other tissues can be integrated. However, heavy-tail distribution and outliers are common in genomics data, which poses challenges to the effectiveness of current transfer learning approaches. In this paper, we study the transfer learning problem under high-dimensional linear models with t-distributed error (Trans-PtLR), which aims to improve the estimation and prediction of target data by borrowing information from useful source data and offering robustness to accommodate complex data with heavy tails and outliers. In the oracle case with known transferable source datasets, a transfer learning algorithm based on penalized maximum likelihood and expectation-maximization algorithm is established. To avoid including non-informative sources, we propose to select the transferable sources based on cross-validation. Extensive simulation experiments as well as an application demonstrate that Trans-PtLR demonstrates robustness and better performance of estimation and prediction when heavy-tail and outliers exist compared to transfer learning for linear regression model with normal error distribution.

Data integration, Variable selection, T distribution, Expectation maximization algorithm, Genotype-Tissue Expression, Cross validation.

and 82373692, QG, NO 82204163; URL: <https://www.nsf.gov.cn/>), Shanghai Rising-Star Program (YY, 21QA1401300; URL: <https://stcsm.sh.gov.cn/xwzx/z/shkjqmd/>), Shanghai Municipal Natural Science Foundation (YY, 22ZR1414900; URL: <https://stcsm.sh.gov.cn/zwgk/zfxgkzbzml/zd gz/jcyj/shzrkxjj/>), and Shanghai Municipal Science and Technology Major Project (GQ, ZD2021CY001; URL: https://service.shanghai.gov.cn/XingZhengWenDangKuJyh/XZGFDDetails.aspx?docid=REPORT_NDOC_008351), and Fundamental Research Program of Shanxi Province (QG, 202203021212382; URL: https://www.shanxi.gov.cn/zfxgk/zfxgkz/zc/xzgfxwj/bmgfxwj1/szfzcbm_76475/skxjst_76478/202404/t20240411_9536094.shtml). All the sponsors or funders played a role in the study design, data collection and analysis, decision to publish, and preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Author summary

Many genetic loci have been shown to be associated with the mechanisms of important disease onset. Therefore, studying the expression of important genes contributes to the diagnosis and treatment of diseases. However, limited target gene expression data poses challenges to studying gene regulation. How to effectively integrate gene expression data from multiple sources is a key issue that needs to be addressed. In this study, we propose a robust transfer learning method aimed at improving the estimation and prediction performance of target gene expression data by integrating information from multiple data sources. By introducing a high-dimensional linear regression model with t-error distribution, our method addresses the shortcomings of previous transfer learning methods faced with heavy-tail distributions and outliers in genomics, providing robustness to complex data features. Extensive simulation experiments and an application demonstrate that our method exhibits better estimation and prediction performance when dealing with gene expression data with heavy-tail distributions and outliers.

1 Introduction

With the rapid development of genomic technologies, there are gene expression data available in public databases for many biological and medical research questions. Integrating multi-source data can overcome the problem of insufficient representativeness of target data, providing us with opportunities to explore the mechanisms of gene regulation and to understand the occurrence and progression of diseases [1,2]. However, due to the heterogeneity of biological samples from different sources, how to integrate the useful information from source data to improve estimation and prediction performance is a key challenge for the analysis and application of gene expression data [3,4].

1.1 Related work

To address the issue of systematic differences in multi-source gene expression data integration, current research mainly employs Meta-analysis (MA) and Data Merging (DM) approaches [5,6]. DM merges all samples in a unique dataset and subsequently performs analyses, while MA aggregates information from different samples at different stages of the analysis, for example, at the beginning by merging all samples or at the end by combining all the results of the separate analysis of specific samples [7,8]. Both methods increase the sample size by pooling all samples but fail to adjust for sample-specific heterogeneity, especially when focusing on the characteristics of specific target data. The heterogeneity among sources not only includes systematic variations caused by experimental conditions and measurement errors but also biological differences, such as variations in gene expression across different human tissue samples [9,10]. A unified model trained on all samples may exhibit bias and be unable to represent the true characteristics of the target sample of interest.

To address the challenge posed by such heterogeneity of source datasets and lack of representation of target data, transfer learning has emerged as a promising approach, aiming to transfer useful information from related but different source datasets to enhance the learning performance of the target data [11, 12]. When the target data is limited and the source data is large but biased, Bouveyron and Jacques (2010) proposed adaptive models to learn the target parameters through a linear transformation of the source model's parameters, showing significant improvements over prediction from the target data only [13]. However, this approach

gives very low adaption freedom due to the limiting of the number of transformation matrix elements learned from the target data to control the extent of transfer learning. Bastani (2021) proposed a more flexible two-step transfer learning estimator for high-dimensional linear models with a single informative auxiliary study [14]. By estimating the estimator from the source data and debiasing it using the target data, significant improvements in prediction performance for target data can be achieved. Furthermore, Li et al. (2022) further considered the estimation and prediction of a high-dimensional linear regression in the setting of two-step transfer learning with multiple source data and proposed a transferable source data detection algorithm to avoid negative transfer [15]. However, the transfer learning estimators proposed by Bastani (2021) and Li et al. (2022) both rely on quadratic optimization based on linear regression with normal distributed error, which is sensitive to heavy-tail and outliers that are common in practice [16, 17].

1.2 Our contributions

In this paper, building on the method proposed by Li et al. (2022) [15], we propose a transfer learning framework under high-dimensional linear models with t -distributed error, aimed at enhancing the robustness of transfer learning under conditions of heavy-tailed distributions and outliers, thereby improving the estimation and prediction of target data. The heavy-tailed characteristics of the t -distribution make it more robust to outliers compared to the normal distribution [18–22]. Different degrees of freedom parameters of the t -distribution can be estimated to adapt to source data with varying outliers or heavy-tailed error distributions [23]. By introducing t -distributed errors, we use the expectation-maximization (EM) algorithm to optimize the penalized likelihood function for estimating regression parameters. In the oracle case with known transferable source datasets, a three-step transfer learning algorithm is established, which combines the information from the target dataset and transferable source datasets, and adjusts their differences simultaneously. To avoid including non-informative sources, we propose to select transferable sources based on cross-validation, which ensures that the data integration will increase the prediction performance.

We not only achieve robustness improvements methodologically but demonstrate that our proposed method can transfer useful information from source datasets to target data and is robust to heavy tails and outliers through extensive simulations and application. Compared to other candidate methods, it exhibits higher accuracy in estimation, prediction, and variable selection.

2 Materials and methods

2.1 Penalized linear regression with t -distributed error

Regression analysis is one of the most widely used statistical methods to understand the relationship between a response and a set of predictors [24]. For subject $i = 1, \dots, n$, let Y_i be a response and $\mathbf{X}_i = (X_i^1, \dots, X_i^p)^T$ be p predictors. The linear regression model can be written as

$$Y_i = \mathbf{X}_i^T \boldsymbol{\beta} + \varepsilon_i, \quad (1)$$

where $\boldsymbol{\beta} = (\beta^1, \dots, \beta^p)^T$ is the coefficient vector of interest and ε_i is the random error. In this paper, we assume that the dimension p is high but $\boldsymbol{\beta}$ is sparse, that is, only a small subset of predictors is relevant to the response. In traditional linear regression, it is assumed that ε_i follows a normal distribution. However, in genomics data, this assumption may not always be met due to outliers or heavy-tailed distributions, then the ordinary least squares estimates may become

biased or inefficient. In this paper, we consider the robust linear regression which assumes that ε_i follow a t -distribution, that is, $\varepsilon_i \sim t(0, \sigma^2, \nu)$ with location 0, scale parameter σ^2 and degrees of freedom ν [15]. The t -distribution has heavier tails than the normal distribution, making it less sensitive to outliers and more suitable for dealing with genomics data that exhibit non-normal errors [18].

The penalized log-likelihood function of $\theta = (\beta^T, \sigma^2, \nu)^T$ for model (1) is given by $\mathcal{L}(\beta, \sigma^2, \nu) - \lambda \|\beta\|_1$, where

$$\mathcal{L}(\beta, \sigma^2, \nu) = \sum_{i=1}^n \log\{t(Y_i; X_i^T \beta, \sigma^2, \nu)\},$$

with

$$t(Y_i; X_i^T \beta, \sigma^2, \nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2}) \sqrt{\pi \nu \sigma^2}} \left(1 + \frac{(Y_i - X_i^T \beta)^2}{\nu \sigma^2}\right)^{-\frac{\nu+1}{2}}, \tag{2}$$

where $\Gamma(\cdot)$ denotes the gamma function, $\lambda \|\beta\|_1$ is the l_1 -penalty term for variable selection [25], and λ is a regularization parameter that determines the amount of shrinkage. The maximum likelihood estimator $\hat{\theta}$ is defined as the maximizer of $\mathcal{L}(\beta, \sigma^2, \nu) - \lambda \|\beta\|_1$. However, directly optimizing the above objective function is computationally complex. In the following, we introduce an alternative representation of [2], which is based on the fact that the t -distribution can be written as a gamma-normal hierarchical form [26] as:

$$Y_i | \tau_i \sim N\left(X_i^T \beta, \frac{\sigma^2}{\tau_i}\right), \tau_i \sim \Gamma\left(\frac{\nu}{2}, \frac{\nu}{2}\right), i = 1, \dots, n, \tag{3}$$

where $N(\cdot)$ denotes the normal distribution and $\Gamma(\cdot, \cdot)$ denotes the gamma distribution. The τ_i is a latent variable, then the gamma-normal hierarchical form leads to expectation-maximization (EM) algorithm [27] implementations for maximum likelihood estimation of the unknown parameters.

Given the data $D = (D_1^T, \dots, D_n^T)$ with $D_i = (X_i^T, Y_i)^T$ for $i = 1, \dots, n$, and the latent variable vector $\tau = (\tau_1, \dots, \tau_n)^T$. According to the gamma-normal hierarchical form (3), the penalized log-likelihood function can be re-expressed as $\mathcal{L}_1(\beta, \sigma^2; D, \tau) + \mathcal{L}_2(\nu; \tau) - \lambda \|\beta\|_1$, where

$$\mathcal{L}_1(\beta, \sigma^2; D, \tau) = \sum_{i=1}^n \left\{ \log \frac{\sqrt{\tau_i}}{\sqrt{2\pi\sigma}} - \frac{\tau_i}{2\sigma^2} (Y_i - X_i^T \beta)^2 \right\},$$

with

$$\mathcal{L}_2(\nu; \tau) = \sum_{i=1}^n \log \frac{\tau_i^{\frac{\nu-2}{2}} e^{-\frac{\nu}{2}\tau_i} \left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)}. \tag{4}$$

In the E-step, given the parameter values $\theta^{(u-1)}$ at the $(u - 1)$ -th iteration, the expectation of the complete log-likelihood is given as

$$Q(\theta, \theta^{(u-1)}) = E\{\mathcal{L}_1(\beta, \sigma^2; D, \tau) | D, \theta^{(u-1)}\} + E\{\mathcal{L}_2(\nu; \tau) | D, \theta^{(u-1)}\} - \lambda \|\beta\|_1.$$

In the M-step, the above objective function can be optimized by Newton-Raphson-type algorithm. The detailed procedure is presented in **Algorithm 1**.

Algorithm 1 EM algorithm

Input: Data D and initial value $\theta^{(0)}$.

At the u -th iteration, given $\theta^{(u-1)}$ from the $(u - 1)$ -th iteration:

E-step. Compute the expectation $E\{\mathcal{L}_1(\beta, \sigma^2; \mathbf{D}, \tau) | \mathbf{D}, \theta^{(u-1)}\}$ and $E\{\mathcal{L}_2(v; \tau) | \mathbf{D}, \theta^{(u-1)}\}$.

M-step. Update unknown parameters as $\theta^{(u)} = \operatorname{argmax}_{\theta} Q(\theta, \theta^{(u-1)})$. Iterate E-step and M-step, until $\|\theta^{(u)} - \theta^{(u-1)}\|_1 < 10^{-6}$.

Output: $\hat{\beta} = \beta^{(u)}$.

2.2 Target dataset, source dataset, and transferable dataset

Suppose there is a target dataset $\{X_i^{(0)}, Y_i^{(0)}\}_{i \in \mathcal{I}^{(0)}}$ and S independent source datasets

$\left\{ \left\{ X_i^{(s)}, Y_i^{(s)} \right\}_{i \in \mathcal{I}^{(s)}} \right\}_{s=1, \dots, S}$, where $\mathcal{I}^{(0)}$ is the index set of subjects belonging to target dataset with

sample size n_0 , and $\mathcal{I}^{(s)}$ ($s = 1, \dots, S$) is the index set of subjects belonging to the s -th source dataset with sample size n_s . The regression models corresponding to the target dataset and the source datasets are:

$$Y_i^{(s)} = X_i^{T(s)} \beta^{(s)} + \varepsilon_i^{(s)}, s = 0, \dots, S, i \in \mathcal{I}^{(s)}. \tag{5}$$

where $\beta^{(0)}$ is the coefficient of the target dataset, and $\beta^{(s)}$ ($s = 1, \dots, S$) is the coefficient of the s -th source dataset.

Our goal is to transfer useful information from the source datasets to the target dataset, to improve the estimation accuracy of $\beta^{(0)}$. The index set of the transferable source datasets is denoted as \mathcal{T} , which is a subset of $\{1, \dots, S\}$. Intuitively, the closer $\beta^{(s)}$ of the s -th source dataset is to $\beta^{(0)}$ of the target dataset, the more transferable that source dataset may be. Deeply, if we combine the s -th source dataset with the target dataset, the estimation or prediction performance is improved relative to using the target dataset alone, then we consider that source dataset to be transferable.

2.3 Transfer learning with known transferable dataset

We first consider that index set \mathcal{T} for the transferable source dataset is known. Motivated by Li et al. (2023) [28], we propose a three-step transfer learning framework for linear regression model [5]. In the first step, we fit regression model using Algorithm 1 in each transferable source dataset $s \in \mathcal{T}$, and obtain $\hat{\beta}^{(s)}$. In the second step, we use the target data to measure the difference between $\beta^{(s)}$ ($s \in \mathcal{T}$) and $\beta^{(0)}$, which is denoted as $\hat{\delta}^{(s)}$. In the third step, we combine the target dataset and transferable source datasets to jointly estimate $\beta^{(TL)}$, and use $\hat{\delta}^{(s)}$ to adjust the differences. The details of the algorithm are shown in **Algorithm 2**, in which the regularization parameters $\lambda_{\beta}^{(s)}$ ($s \in \mathcal{T}$), λ_{δ} , and λ_{β} are selected by cross validation.

Algorithm 2 Transfer learning with known transferable dataset \mathcal{T}

Input: Target dataset $\{D_i = (X_i^T, Y_i)^T\}_{i \in \mathcal{I}^{(0)}}$, source datasets

$\{D_i = (X_i^T, Y_i)^T\}_{i \in \mathcal{I}^{(s)}}$ for $s \in \mathcal{T}$, and index set \mathcal{T} for the transferable dataset.

Step 1. Fit the regression model in each transferable source dataset. For $s \in \mathcal{T}$, compute

$$\hat{\beta}^{(s)} = \operatorname{argmax}_{\beta} \left\{ \mathcal{L}(\beta; \{D_i\}_{i \in \mathcal{I}^{(s)}}) - \lambda_{\beta}^{(s)} \|\beta\|_1 \right\}.$$

Step 2. Measure the differences using target dataset. For $s \in \mathcal{T}$, compute

$$\widehat{\boldsymbol{\delta}}^{(s)} = \operatorname{argmax}_{\boldsymbol{\delta}} \left\{ \mathcal{L}(\widehat{\boldsymbol{\beta}}^{(s)} + \boldsymbol{\delta}; \{\mathbf{D}_i\}_{i \in \mathcal{I}^{(0)}}) - \lambda_{\boldsymbol{\delta}}^{(s)} \|\boldsymbol{\delta}\|_1 \right\}.$$

Step 3: Joint estimation using target dataset and transferable source datasets, and adjusting the differences. Compute

$$\widehat{\boldsymbol{\beta}}^{(TL)} = \operatorname{argmax}_{\boldsymbol{\beta}} \left\{ \mathcal{L}(\boldsymbol{\beta}; \{\mathbf{D}_i\}_{i \in \mathcal{I}^{(0)}}) + \sum_{s \in \mathcal{T}} \mathcal{L}(\boldsymbol{\beta} - \widehat{\boldsymbol{\delta}}^{(s)}; \{\mathbf{D}_i\}_{i \in \mathcal{I}^{(s)}}) - \lambda_{\boldsymbol{\beta}} \|\boldsymbol{\beta}\|_1 \right\}.$$

The above optimization problems can be achieved by running EM **Algorithm 1**.

Output: $\widehat{\boldsymbol{\beta}}^{(TL)}$.

2.4 Detecting the transferable dataset

We now consider how to detect the index set \mathcal{T} for the transferable dataset. As previously mentioned, if we combine the s -th source dataset with the target dataset, the estimation or prediction performance is improved relative to using the target dataset alone, then we consider that source dataset to be transferable. Inspired by Eaton et al. (2008) and Zhang and Zhu (2022) [29, 30], we propose a cross-validation framework for detecting the transferable dataset. We randomly split the subjects in target dataset into two disjoint equal-size groups $\mathcal{I}^{(0)} = \mathcal{I}^{(0, \text{tr})} \cup \mathcal{I}^{(0, \text{va})}$, where $\mathcal{I}^{(0, \text{tr})}$ is the training set and $\mathcal{I}^{(0, \text{va})}$ is the validation set. We first fit regression model using Algorithm 1 in training set $\mathcal{I}^{(0, \text{tr})}$, and obtain $\widehat{\boldsymbol{\beta}}^{(0)}$. For a given $s = 1, \dots, S$, then we perform transfer learning using Algorithm 1 in $\mathcal{I}^{(0, \text{tr})}$ with the s -th source dataset, and obtain $\widehat{\boldsymbol{\beta}}^{(s)}$. We evaluate the performance of $\widehat{\boldsymbol{\beta}}^{(0)}$ and $\widehat{\boldsymbol{\beta}}^{(s)}$ on the validation set, by computing $\widehat{T}^{(s)} = \mathcal{L}(\widehat{\boldsymbol{\beta}}^{(0)}; \{\mathbf{D}_i\}_{i \in \mathcal{I}^{(0, \text{va})}}) - \mathcal{L}(\widehat{\boldsymbol{\beta}}^{(s)}; \{\mathbf{D}_i\}_{i \in \mathcal{I}^{(0, \text{va})}})$. The $\widehat{T}^{(s)} < 0$ indicates that the transfer of the s -th source dataset makes the prediction performance increase, so we consider $s \in \mathcal{T}$. If $\widehat{T}^{(s)} > 0$ but not exceeding $\boldsymbol{\epsilon} \left(\mathcal{L}(\widehat{\boldsymbol{\beta}}^{(0)}; \{\mathbf{D}_i\}_{i \in \mathcal{I}^{(0, \text{va})}}) \vee 0.01 \right)$ for a given threshold $\boldsymbol{\epsilon}$, we also consider $s \in \mathcal{T}$. The details of the algorithm are shown in **Algorithm 3**.

Algorithm 3 Transfer learning with unknown transferable dataset \mathcal{T}

Input: Target dataset $\{\mathbf{D}_i = (\mathbf{X}_i^T, Y_i)^T\}_{i \in \mathcal{I}^{(0)}}$, source dataset

$$\{\mathbf{D}_i = (\mathbf{X}_i^T, Y_i)^T\}_{i \in \bigcup_{s=1}^S \mathcal{I}^{(s)}}.$$

1. Detecting transferable dataset

Randomly split $\mathcal{I}^{(0)}$ into two disjoint equal-size groups $\mathcal{I}^{(0)} = \mathcal{I}^{(0, \text{tr})} \cup \mathcal{I}^{(0, \text{va})}$.

(1) Compute $\widehat{\boldsymbol{\beta}}^{(0)} = \operatorname{argmax}_{\boldsymbol{\beta}} \left\{ \mathcal{L}(\boldsymbol{\beta}; \{\mathbf{D}_i\}_{i \in \mathcal{I}^{(0, \text{tr})}}) - \lambda_{\boldsymbol{\beta}}^{(0)} \|\boldsymbol{\beta}\|_1 \right\}$.

(2) For $s = 1, \dots, S$, compute $\widehat{\boldsymbol{\beta}}^{(s)}$ by running the **Algorithm 2** with $\{\mathbf{D}_i\}_{i \in \mathcal{I}^{(0, \text{tr})} \cup \mathcal{I}^{(s)}}$.

(3) For $s = 1, \dots, S$, compute $\widehat{T}^{(s)} = \mathcal{L}(\widehat{\boldsymbol{\beta}}^{(0)}; \{\mathbf{D}_i\}_{i \in \mathcal{I}^{(0, \text{va})}}) - \mathcal{L}(\widehat{\boldsymbol{\beta}}^{(s)}; \{\mathbf{D}_i\}_{i \in \mathcal{I}^{(0, \text{va})}})$.

Let $\widehat{\mathcal{T}} = \left\{ s : \widehat{T}^{(s)} < \boldsymbol{\epsilon} \left(\mathcal{L}(\widehat{\boldsymbol{\beta}}^{(0)}; \{\mathbf{D}_i\}_{i \in \mathcal{I}^{(0, \text{va})}}) \vee 0.01 \right) \right\}$.

2. Run **Algorithm 2** with target dataset $\{\mathbf{D}_i\}_{i \in \mathcal{I}^{(0)}}$, and detected transferable source dataset $\{\mathbf{D}_i\}_{i \in \mathcal{I}^{(s)}}$ for $s \in \widehat{\mathcal{T}}$ to obtain $\widehat{\boldsymbol{\beta}}^{(TL)}$.

Output: $\widehat{\boldsymbol{\beta}}^{(TL)}$

3. Numerical experiments

In this section, we conduct simulation studies to evaluate the performance of the proposed method. We compare our proposed transfer learning framework for penalized t -linear regression (Trans-PtLR based on Algorithm 3) with (i) Naive Trans-PtLR: based on Algorithm 2 using all source datasets; (ii) PtLR: based on Algorithm 1 using only target dataset. We also compare them with the following three methods, Trans-PNLR, Naive Trans-PNLR, and PNLR, which consider the penalized normal-linear regression.

3.1 Simulation

We consider a target dataset with $n_0 = 150$, and $S = 10$ source datasets with $n_1 = \dots n_S = 100$. For $s = 0, \dots, S$, we set $p = 500$ and the predictors $X_i^{(s)}$ is drawn from the multivariate normal distribution with mean zero and covariance matrix $\Sigma(\rho)$, where $\Sigma(\rho)$ is the first-order autoregressive correlation structure with $\rho = 0.7$. The response is generated as

$$Y_i^{(s)} = X_i^{T(s)} \beta^{(s)} + \epsilon_i^{(s)}.$$

For the coefficient $\beta^{(0)} = (\beta^{1(0)}, \dots, \beta^{p(0)})^T$ corresponding to target dataset, we set $k = 16$, $\beta^{j(0)} = 0.5$ for $j = 1, \dots, k$ and $\beta^{j(0)} = 0$ otherwise. We randomly select $|\mathcal{T}| \in \{0, 2, 4, 6, 8, 10\}$ source datasets as the transferable datasets with the corresponding coefficients set to be

$$\beta^{j(s)} = \begin{cases} \beta^{j(0)} - \frac{h}{100} I\{j \in H^{(s)}\} & s \in \mathcal{T}, \\ \frac{h}{100} I\{j \in H^{(s)} \cup [k]\} & s \notin \mathcal{T}, \end{cases}$$

where $H^{(s)}$ is a random subset of $\{17, \dots, p\}$ with $|H^{(s)}| = 50$ and $h \in \{10, 20\}$ controls the heterogeneity levels between target dataset and source datasets.

To evaluate the robustness of the proposed method, we consider four distributions of the error. For $s = 0, \dots, S$, we set (i) normal distribution as $\epsilon^{(s)} \sim N(0, 1)$; (ii) t -distribution as $\epsilon^{(s)} \sim t(0, 1, 5)$; (iii) contaminated normal distribution as $\epsilon^{(s)} \sim 0.9N(0, 1) + 0.1N(0, 10)$, which generates outliers; (iv) skew t -distribution as $\epsilon^{(s)} \sim St(0, 1, 5, \text{skewness} = 1)$.

We conduct 100 replications. In each replication, to evaluate the estimation accuracy, we calculate the estimation error $\|\widehat{\beta}^{(0)} - \beta^{(0)}\|_2^2$; to evaluate the prediction accuracy, we randomly split the target dataset into five folds, using four folds to train and the remaining fold to calculate the mean prediction error $(\widehat{Y}_i - Y_i)^2$ with $\widehat{Y}_i = X_i^{T(0)} \widehat{\beta}^{(0)}$; to evaluate the accuracy of the variable selection, we calculate the proportion of true predictors among predictors selected by l_1 -regularization (defined as precision) and the proportion of selected true predictors among all true predictors (defined as recall).

The threshold ϵ in **Algorithm 3** is set to be 0.1. Results corresponding to $\epsilon = 0.05$ or 0.15 are available in **S1–S4 Fig**, which shows that the estimates are insensitive to a small range of ϵ .

3.2 Application to the prediction of gene expression

Gene expression data can be used to explore the gene-gene interaction and gene regulation for a better understanding of molecular processes of pathogenesis and the purpose of prediction of disease. For instance, discovering gene expression patterns that are characteristic of a certain disease to differentiate between healthy and diseased individuals.

In this section, we consider Primary Familial Brain Calcification (PFBC) which is a hereditary neurodegenerative disease characterized by progressive bilateral cerebral calcifications,

accompanied by various symptoms such as dystonia, ataxia, Parkinson's disease, dementia, depression, headaches, and epilepsy [31]. Currently, the exact etiology of PFBC remains unclear. Recent studies indicate that variations in the JAM2 gene result in decreased JAM2 mRNA expression and loss of JAM2 protein in patient fibroblasts, consistent with the loss-of-function mechanism [32, 33]. JAM2 is a protein-coding gene located on chromosome 21, and the encoded protein is a cell adhesion molecule expressed in various tissues and cell types, influencing cell migration, adhesion, and interactions [34]. The expression pattern of JAM2 may be regulated under physiological and pathological conditions. Therefore, obtaining the expression levels of JAM2 is of great significance for the diagnosis of PFBC. Given the technical challenges and ethical concerns associated with brain tissue sequencing, constructing model to predict the expression levels of JAM2 in target brain tissues holds potential clinical diagnostic value for PFBC. The objective of this application is to train a prediction model of JAM2 expression levels in target brain tissues, transferring useful information from other tissues or cell types to improve the prediction of JAM2 expression levels in the target brain tissue.

We evaluate the predictive performance of our proposed algorithm using tissue-specific gene expression data from the Genotype-Tissue Expression (GTEx) version 8 dataset, which consists of 838 donors and 17,382 samples from 54 non-diseased tissue types (S1 Table). In the analysis of this study, we used 49 tissues or cell lines that had at least 70 individuals, including a total of 17,329 samples from 838 donors. Based on previously published literature and publicly available lists of central nervous system (CNS)-related genes, the CNS-related genes were assembled as MODULE 137, including 546 genes as well as 1,632 additional genes that are significantly enriched in the same experiments as the genes of the module. These genes may participate in similar biological pathways or regulatory networks, working together in the central nervous system to perform specific functions. Therefore, they provide an important foundation for our study of JAM2 genes in PFBC. We consider 13 brain tissues as target tissues and the remaining 36 tissues as source tissues. The average sample size of target tissues and source tissues was 203 and 408, respectively. After excluding missing values, the final predictors included 1292 genes.

To compare the performance of the proposed method, we conduct Trans-PNLR and Trans-PtLR models in each tissue based on the association of the JAM2 gene and other CNS-related genes and identify the informative tissues to transfer to improve the performance of the target model in each brain tissue. We also conduct Naive Trans-PNLR and Naive Trans-PtLR to understand the total information level of all source tissues. We also conduct PNLR and PtLR to evaluate the prediction performance using only the target tissue, without transforming information from other source tissues. The response variable is the JAM2 expression level and covariates are the expression levels of other genes included in MODULE 137. For each model, we randomly split the target dataset into five folds, using four folds to train and the remaining fold to calculate the mean prediction error $\left(\widehat{Y}_i - Y_i\right)^2$ with $\widehat{Y}_i = \mathbf{X}_i^{T(0)}\widehat{\boldsymbol{\beta}}^{(0)}$. The data is standardized before analysis.

4 Results

4.1 Simulation results

Figs 1 and 2 and Table 1 show the estimation error, relative prediction error, and variable selection precision and recall of candidate methods under different heterogeneity levels ($h = 10$ or 20), different numbers of transferable source datasets ($|\mathcal{T}| \in \{0, 2, 4, 6, 8, 10\}$), and different distributions of the error ($\epsilon^{(s)}$ follows normal distribution (N), t -distribution (t), contaminated normal distribution (CN), and skew t -distribution (St)).

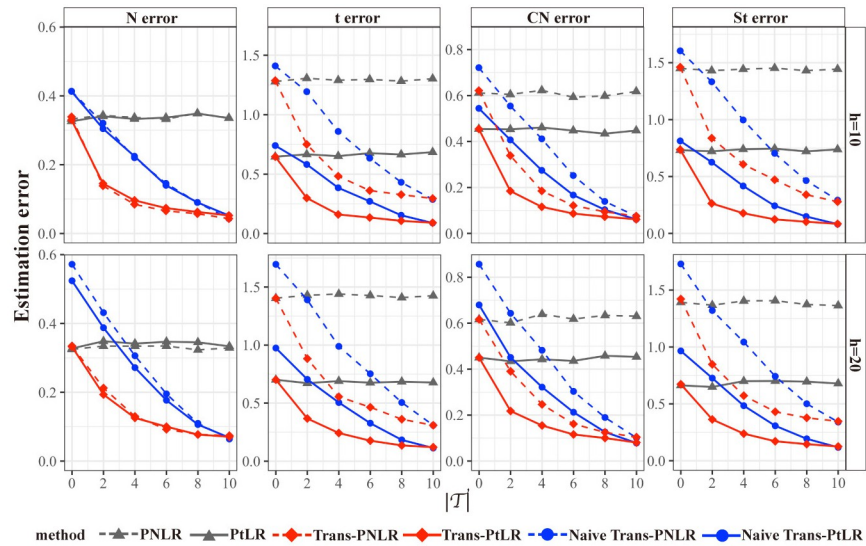


Fig 1. Estimation error $\|\hat{\beta}^{(0)} - \beta^{(0)}\|_2^2$ of candidate methods with a threshold ϵ of 0.1. The estimation error is evaluated under different heterogeneity levels ($h = 10$ or 20), different numbers of transferable source datasets ($|\mathcal{T}| \in \{0, 2, 4, 6, 8, 10\}$), and different distributions of the error ($\epsilon^{(s)}$ follows normal distribution (N), t -distribution (t), contaminated normal distribution (CN), and skew t -distribution (St)).

<https://doi.org/10.1371/journal.pcbi.1012739.g001>

For the normal distributed (N) error, Trans-PNLR and Trans-PtLR perform better than PNLR and PtLR, which indicates that transfer learning can help us extract additional information from the source datasets to improve the accuracy of estimation and prediction on the target dataset. As the numbers of transferable datasets $|\mathcal{T}|$ increases, the accuracy of estimation

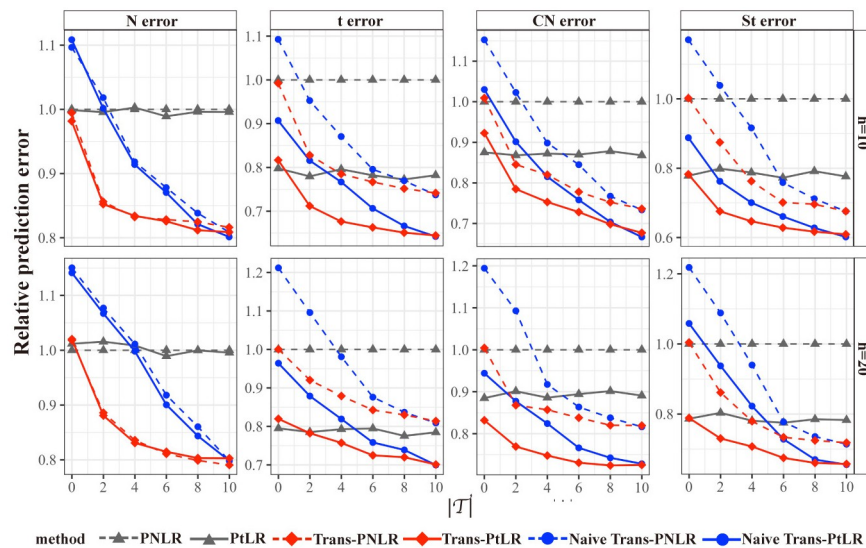


Fig 2. Relative prediction error of candidate methods relative to the PNLR with a threshold ϵ of 0.1. The relative prediction error is evaluated under different heterogeneity levels ($h = 10$ or 20), different numbers of transferable source datasets ($|\mathcal{T}| \in \{0, 2, 4, 6, 8, 10\}$), and different distributions of the error ($\epsilon^{(s)}$ follows normal distribution (N), t -distribution (t), contaminated normal distribution (CN), and skew t -distribution (St)). We randomly split the target dataset into five folds, using four folds to train and the remaining fold to calculate the mean prediction error $(\hat{Y}_i - Y_i)^2$ with $\hat{Y}_i = \mathbf{X}_i^{T(0)} \hat{\beta}^{(0)}$.

<https://doi.org/10.1371/journal.pcbi.1012739.g002>

Table 1. Precision and recall (in parentheses) of candidate methods under different heterogeneity levels ($h = 10$ or 20), different numbers of transferable source datasets ($|\mathcal{T}| \in \{0, 2, 4, 6, 8, 10\}$), and different distributions of the error ($\epsilon^{(s)}$ follows normal distribution (N), t -distribution (t), contaminated normal distribution (CN), and skew t -distribution (St)).

| | N error | | t error | | CN error | | St error | |
|----------------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| | PNLR | PtLR | PNLR | PtLR | PNLR | PtLR | PNLR | PtLR |
| $h = 10$ | | | | | | | | |
| No transfer | 0.71 (1.00) | 0.72 (1.00) | 0.71 (0.92) | 0.73 (0.99) | 0.71 (0.99) | 0.72 (1.00) | 0.71 (0.92) | 0.74 (0.99) |
| Transfer | | | | | | | | |
| $ \mathcal{T} = 0$ | 0.69 (1.00) | 0.71 (1.00) | 0.68 (0.93) | 0.72 (0.99) | 0.67 (0.99) | 0.69 (1.00) | 0.68 (0.93) | 0.72 (0.98) |
| $ \mathcal{T} = 2$ | 0.76 (1.00) | 0.86 (1.00) | 0.65 (0.99) | 0.72 (1.00) | 0.62 (1.00) | 0.77 (1.00) | 0.65 (0.99) | 0.73 (1.00) |
| $ \mathcal{T} = 4$ | 0.82 (1.00) | 0.95 (1.00) | 0.64 (1.00) | 0.78 (1.00) | 0.67 (1.00) | 0.89 (1.00) | 0.65 (1.00) | 0.75 (1.00) |
| $ \mathcal{T} = 6$ | 0.90 (1.00) | 0.97 (1.00) | 0.67 (1.00) | 0.86 (1.00) | 0.75 (1.00) | 0.96 (1.00) | 0.69 (1.00) | 0.85 (1.00) |
| $ \mathcal{T} = 8$ | 0.94 (1.00) | 0.99 (1.00) | 0.74 (1.00) | 0.92 (1.00) | 0.82 (1.00) | 0.97 (1.00) | 0.71 (1.00) | 0.91 (1.00) |
| $ \mathcal{T} = 10$ | 0.96 (1.00) | 0.99 (1.00) | 0.78 (1.00) | 0.95 (1.00) | 0.88 (1.00) | 0.97 (1.00) | 0.75 (1.00) | 0.95 (1.00) |
| $h = 20$ | | | | | | | | |
| No transfer | 0.71 (1.00) | 0.72 (1.00) | 0.71 (0.91) | 0.73 (0.99) | 0.71 (0.99) | 0.71 (1.00) | 0.70 (0.90) | 0.73 (0.99) |
| Transfer | | | | | | | | |
| $ \mathcal{T} = 0$ | 0.70 (1.00) | 0.72 (1.00) | 0.68 (0.92) | 0.73 (0.98) | 0.70 (0.99) | 0.72 (1.00) | 0.68 (0.91) | 0.71 (0.99) |
| $ \mathcal{T} = 2$ | 0.65 (1.00) | 0.79 (1.00) | 0.64 (0.99) | 0.76 (1.00) | 0.68 (1.00) | 0.76 (1.00) | 0.65 (0.98) | 0.68 (1.00) |
| $ \mathcal{T} = 4$ | 0.73 (1.00) | 0.85 (1.00) | 0.63 (0.99) | 0.79 (1.00) | 0.66 (1.00) | 0.81 (1.00) | 0.68 (1.00) | 0.77 (1.00) |
| $ \mathcal{T} = 6$ | 0.78 (1.00) | 0.92 (1.00) | 0.63 (1.00) | 0.86 (1.00) | 0.70 (1.00) | 0.85 (1.00) | 0.69 (1.00) | 0.85 (1.00) |
| $ \mathcal{T} = 8$ | 0.82 (1.00) | 0.96 (1.00) | 0.68 (1.00) | 0.84 (1.00) | 0.73 (1.00) | 0.91 (1.00) | 0.72 (1.00) | 0.90 (1.00) |
| $ \mathcal{T} = 10$ | 0.86 (1.00) | 0.97 (1.00) | 0.66 (1.00) | 0.89 (1.00) | 0.74 (1.00) | 0.94 (1.00) | 0.74 (1.00) | 0.93 (1.00) |

Precision: the proportion of true predictors among predictors selected by l_1 -regularization. Recall (in parentheses): the proportion of selected true predictors among all true predictors.

<https://doi.org/10.1371/journal.pcbi.1012739.t001>

and prediction becomes higher. Trans-PNLR and Trans-PtLR also perform better than Naive Trans-PNLR and Naive Trans-PtLR, especially when $|\mathcal{T}|$ is relatively small. The probable reason is that when the number of transferable source datasets is small, transferring all the source datasets will result in negative transfer. Therefore, it is necessary to adopt the Algorithm 3 to detect the transferable dataset (S2 Table). We also observe that when the heterogeneity levels between target dataset and source datasets are smaller ($h = 10$), the accuracy of estimation and prediction for Trans-PNLR and Trans-PtLR is higher.

For the comparison between Trans-PNLR and Trans-PtLR, when the error is generated from normal distribution that contains no outliers, the performances of these two methods are quite similar. When the error is generated from t -distribution with heavy tail, Trans-PtLR has a smaller estimation error and prediction error than Trans-PNLR. When the error is generated from contaminated normal distributions, Trans-PtLR is less sensitive to potential outliers than Trans-PNLR. When the error is generated from skewed t -distribution which is skewed and heavy-tailed, Trans-PtLR also achieved better estimation and prediction performance than Trans-PNLR. As for genomics data, outliers and non-normal errors are likely to be encountered, so our proposed Trans-PtLR may be more suitable for application. For single regression coefficients with non-zero true effect, the estimation error exhibits the similar trend as Fig 1, while for regression coefficients with zero true effect, the estimation errors approach zero due to being compressed to zero by the l_1 -penalty (see S5 and S6 Figs). We also explored the impact of estimation performance when sample size of target and sources vary. After correctly identifying transferable datasets, even if the sample size of the sources is small, the estimation performance will still not be worse than using only the target data (S7 Fig).

We can find from Table 1 that the recall of Trans-PtLR and Trans-PNLR is close to 1, but in some scenarios, such as when $\varepsilon^{(s)}$ follows a t -distribution (t) or skew t -distribution (St), for $h = 10$ or 20 and $|T| = 0$, the recall of Trans-PtLR is higher than that of Trans-PNLR. This indicates that our method almost never misses important predictors. We also found that in most cases, the Precision of Trans-PtLR is higher than that of Trans-PNLR. This suggests that while Trans-PNLR may select more irrelevant variables, Trans-PtLR more accurately excludes unimportant predictors, thereby enhancing precision.

4.2 Predicting the gene expression

Before the analysis, we first fit a normal linear regression model for all tissues and conducted model diagnoses, then, we took several target tissues and source tissues as examples, plotted the distribution of error and compared it to the normal distribution. As shown in Fig 3, the distribution of error is heavy-tailed or may contain outliers, indicating that fitting the tissue-specific gene expression data with a normal linear regression may not be appropriate. It is reasonable to assume the error term following t distributions to reduce the effects of the potential outliers. The transferable sources detected by all methods and the sample size of all transferable sources are available in S3 and S4 Tables.

Fig 4 shows the relative prediction error of five models (PtLR, Naive Trans-PNLR, Naive Trans-PtLR, Trans-PNLR, and Trans-PtLR) relative to the PNL using CNS gene expression

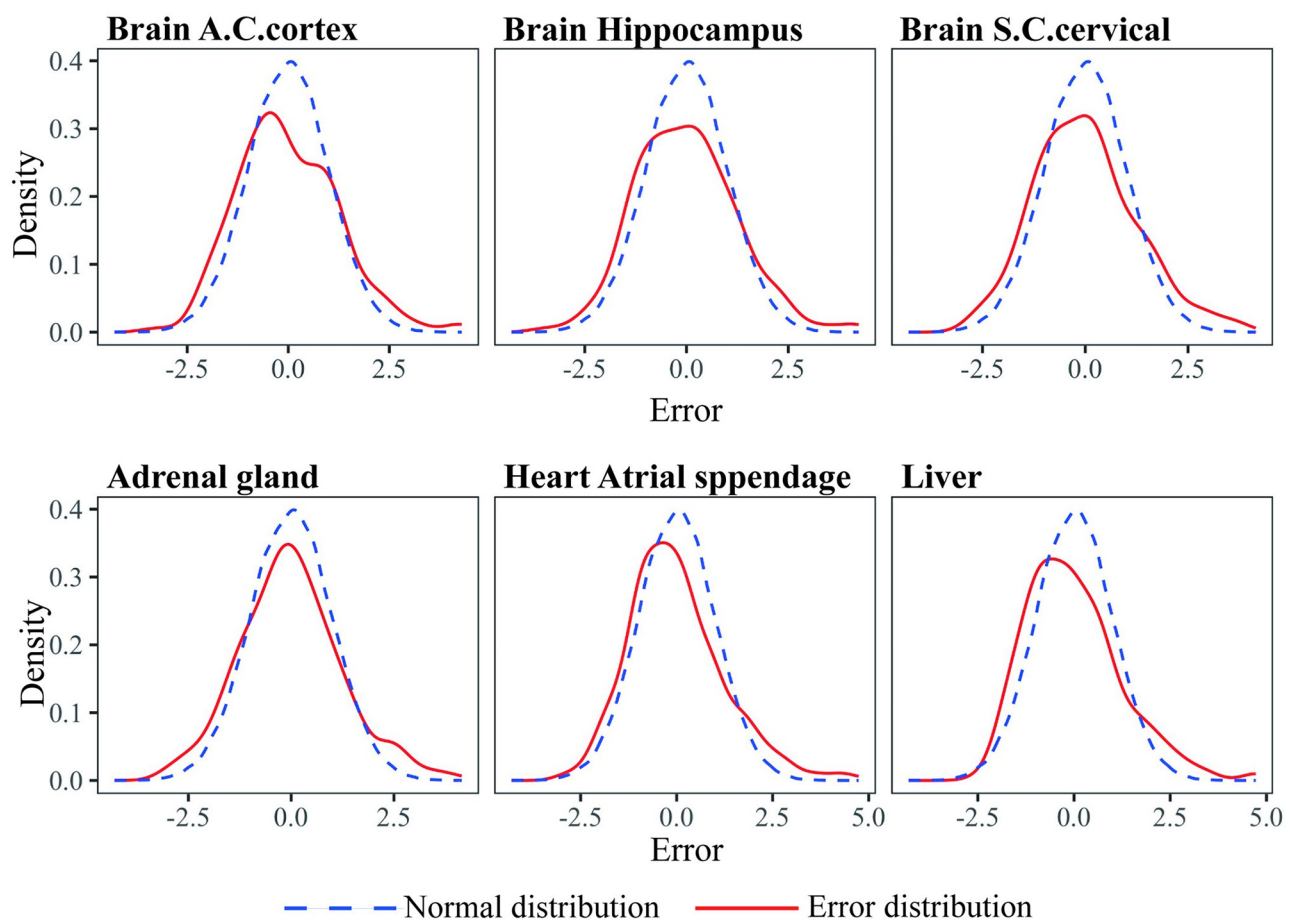


Fig 3. Distribution of error by normal linear regression for six tissues compared with normal distribution.

<https://doi.org/10.1371/journal.pcbi.1012739.g003>

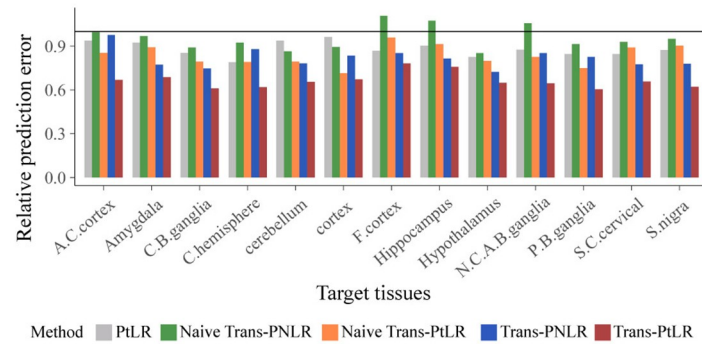


Fig 4. Relative prediction error of PtLR, Naive Trans-PNLr, Naive Trans-PtLR, Trans-PNLr, and Trans-PtLR relative to the PNLr. We randomly split the data in target tissues into five folds, using four folds to train and the remaining fold to calculate the mean prediction error $\left(\widehat{Y}_i - Y_i\right)^2$ with $\widehat{Y}_i = \mathbf{X}_i^{T(0)} \widehat{\beta}^{(0)}$.

<https://doi.org/10.1371/journal.pcbi.1012739.g004>

levels to predict JAM2 gene expression levels in 13 brain tissues. The results show that transfer learning methods significantly reduce prediction errors in most scenarios. This highlights the practical advantages of integrating useful information from source tissues.

It is noteworthy that the model's performance is poorer when transferable source datasets are not pre-identified (as in Naive Trans PNLr and Naive Trans PtLR). This emphasizes the importance of using the proposed transferable source detection algorithm, which effectively identifies valuable source information and improves predictive performance, demonstrating its practical applicability in complex data integration tasks.

Furthermore, the results suggest that transfer learning within the t -linear regression (Trans-PtLR) framework outperforms the normal regression (Trans-PNLr) framework. Transfer learning through Trans-PNLr reduced the average prediction error by 18.3%, while Trans-PtLR further reduced this by 33.6% (with Trans-PtLR averaging a 15.3% reduction in prediction error compared to Trans-PNLr). The Wilcoxon signed-rank test shows that the difference in prediction errors reduced by these two methods has a P -value of 0.0002. This significant improvement indicates that modeling with heavy-tailed t -distributed errors can better handle outliers and enhance the robustness of real-world datasets, particularly in gene expression data where heavy tails and noise are common.

5 Discussion

This paper studies robust transfer learning for penalized t -linear regression model under high-dimensional data and its application in gene expression prediction. We conduct a three-step transfer learning algorithm to obtain the joint estimator and use a data-driven transferable source detection algorithm to prevent negative transfer. Extensive simulations demonstrate the better performance and robustness of estimation and prediction compared with transfer learning for normal linear regression when outliers or heavy-tailed errors exist. In this application based on GTEx data, we build an expression prediction model for the JAM2 gene which treats 1292 other genes as predictors. We consider 13 brain tissues as target tissues and the remaining 36 tissues as source tissues. Our proposed method also shows higher accuracy for prediction compared to another candidate approach in each target tissue.

One of the key distinctions between our Trans-PtLR method and existing methods Trans-PNLr lies in the handling of error distributions. Unlike Trans-PNLr, which relies on the assumption of normal distribution, our method incorporates t -distributed errors, making it

more robust in the presence of outliers and heavy-tailed distributions. This improvement enables Trans-PtLR to more effectively capture true information in complex biological data, thereby enhancing estimation and prediction performance. By providing more accurate gene expression predictions, it aids in the understanding of gene regulatory mechanisms, supporting the advancement of personalized medicine. Furthermore, although the motivation for our research is gene expression prediction, the flexibility and robustness of the proposed framework for robust transfer learning make it potentially applicable to other fields, such as medical image analysis.

However, our method also has some limitations. First, the advantages of this method may be more pronounced under specific conditions, such as in datasets exhibiting skewed or heavy-tailed distributions, as demonstrated by our simulation studies. In contrast, the Trans-PtLR method does not show better performance than the Trans-PNLR method when errors follow a normal distribution. Secondly, our focus in this study is only on linear regression with t -distributed error. How to extend the robust transfer learning framework to other models such as Cox model and quantile regression to apply in the integration of multi-omics data is also an interesting problem. Thirdly, our methods used l_1 -penalty to control model sparsity and information level transferred from source data. It may be extended to l_2 -penalty or elastic net type of penalties, depending on the characteristics of the data at hand and whether the difference of regression parameters between a source and the target data is sparse or nearly sparse [35]. We can further consider different penalties for model sparsity control and transferred information level control. In addition, a prerequisite of our method is that it requires the same predictor set to be available in all sources, which may not be satisfied in real applications. How to deal with missing features appropriately while providing effective information for transfer is a direction that can be considered in the future.

In conclusion, Trans-PtLR is a robust transfer learning approach that can transfer valuable information from multiple source datasets to improve the performance of estimation and prediction in the target dataset under high dimensional scenarios. It offers robustness and flexibility to accommodate complex data with outliers and heavy tails.

Supporting information

S1 Table. The list of 54 tissues in GTEx dataset corresponding to their sample sizes.

(DOCX)

S2 Table. The accuracy of the transferable source detection algorithm in correctly identifying transferable datasets.

(DOCX)

S3 Table. The transferable sources detected by the proposed method Trans-PtLR and sample size of all transferable sources.

(DOCX)

S4 Table. The transferable sources detected by the method Trans-PNLR and sample size of all transferable sources.

(DOCX)

S1 Fig. Estimation error $\|\hat{\beta}^{(0)} - \beta^{(0)}\|_2^2$ of candidate methods with a threshold τ of 0.05.

(TIF)

S2 Fig. Estimation error $\|\hat{\beta}^{(0)} - \beta^{(0)}\|_2^2$ of candidate methods with a threshold τ of 0.15.

(TIF)

S3 Fig. Relative prediction error of candidate methods relative to the PNLN with a threshold ϵ of 0.05. We randomly split the target dataset into five folds, using four folds to train and the remaining fold to calculate the mean prediction error $(\hat{Y}_i - Y_i)^2$ with $\hat{Y}_i = \mathbf{X}_i^{T(0)} \hat{\boldsymbol{\beta}}^{(0)}$.

(TIF)

S4 Fig. Relative prediction error of candidate methods relative to the PNLN with a threshold ϵ of 0.15. We randomly split the target dataset into five folds, using four folds to train and the remaining fold to calculate the mean prediction error $(\hat{Y}_i - Y_i)^2$ with $\hat{Y}_i = \mathbf{X}_i^{T(0)} \hat{\boldsymbol{\beta}}^{(0)}$.

(TIF)

S5 Fig. Estimation error $\|\hat{\beta}_1^{(0)} - \beta_1^{(0)}\|_2^2$ of candidate methods with a threshold ϵ of 0.1.

(TIF)

S6 Fig. Estimation error $\|\hat{\beta}_{20}^{(0)} - \beta_{20}^{(0)}\|_2^2$ of candidate methods with a threshold ϵ of 0.1.

(TIF)

S7 Fig. Estimation error $\|\hat{\beta}^{(0)} - \beta^{(0)}\|_2^2$ of candidate methods with a threshold ϵ of 0.1, showing how estimation error varies with sample size.

(TIF)

S1 Code. File includes code "Transfer learning.R" and "updatebeta.cpp" for implementing transfer learning methods.

(ZIP)

Author Contributions

Conceptualization: Lulu Pan, Qian Gao, Kecheng Wei, Guoyou Qin.

Formal analysis: Lulu Pan, Kecheng Wei.

Funding acquisition: Qian Gao, Yongfu Yu, Guoyou Qin, Tong Wang.

Methodology: Lulu Pan, Kecheng Wei, Guoyou Qin, Tong Wang.

Resources: Yongfu Yu, Guoyou Qin, Tong Wang.

Validation: Lulu Pan.

Visualization: Lulu Pan, Qian Gao.

Writing – original draft: Lulu Pan, Kecheng Wei.

Writing – review & editing: Qian Gao, Yongfu Yu, Guoyou Qin, Tong Wang.

Reference

1. Deng L, Liu D, Li Y, Wang R, Liu J, Zhang J, et al. MSPCD: predicting circRNA-disease associations via integrating multi-source data and hierarchical neural network. *BMC bioinformatics*. 2022; 23(Suppl 3):427. <https://doi.org/10.1186/s12859-022-04976-5> PMID: 36241972
2. Adamer MF, Brüningk SC, Tejada-Arranz A, Estermann F, Basler M, Borgwardt K. reComBat: batch-effect removal in large-scale multi-source gene-expression data integration. *Bioinformatics Advances*. 2022; 2(1):vbac071. <https://doi.org/10.1093/bioadv/vbac071> PMID: 36699372
3. Hamid JS, Hu P, Roslin NM, Ling V, Greenwood CM, Beyene J. Data integration in genetics and genomics: methods and challenges. *Human genomics and proteomics: HGP*. 2009; 2009. <https://doi.org/10.4061/2009/869093> PMID: 20948564
4. Gomez-Cabrero D, Abugessaisa I, Maier D, Teschendorff A, Merckenschlager M, Gisel A, et al. Data integration in the era of omics: current and future challenges. *BMC systems biology*. 2014; 8:1–10.

5. Ramasamy A, Mondry A, Holmes CC, Altman DG. Key issues in conducting a meta-analysis of gene expression microarray datasets. *PLoS Med.* 2008; 5(9):e184. <https://doi.org/10.1371/journal.pmed.0050184> PMID: 18767902
6. Warnat P, Eils R, Brors B. Cross-platform analysis of cancer microarray data improves gene expression based classification of phenotypes. *BMC Bioinformatics.* 2005; 6:265. <https://doi.org/10.1186/1471-2105-6-265> PMID: 16271137
7. Lagani V, Karozou AD, Gomez-Cabrero D, Silberberg G, Tsamardinos I. A comparative evaluation of data-merging and meta-analysis methods for reconstructing gene-gene interactions. *BMC Bioinformatics.* 2016; 17 Suppl 5(Suppl 5):194. <https://doi.org/10.1186/s12859-016-1038-1> PMID: 27294826
8. Lazar C, Meganck S, Taminau J, Steenhoff D, Coletta A, Molter C, et al. Batch effect removal methods for microarray gene expression data integration: a survey. *Brief Bioinform.* 2013; 14(4):469–90. <https://doi.org/10.1093/bib/bbs037> PMID: 22851511
9. Sturm G, List M, Zhang JD. Tissue heterogeneity is prevalent in gene expression studies. *NAR Genomics and Bioinformatics.* 2021; 3(3):lqab077. <https://doi.org/10.1093/nargab/lqab077> PMID: 34514392
10. Zhang JD, Hatje K, Sturm G, Broger C, Ebeling M, Burtin M, et al. Detect tissue heterogeneity in gene expression data with BioQC. *BMC genomics.* 2017; 18:1–9.
11. Weiss K, Khoshgoftaar TM, Wang D. A survey of transfer learning. *Journal of Big data.* 2016; 3(1):1–40.
12. Torrey L, Shavlik J. Transfer learning. *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques: IGI global;* 2010. p. 242–64.
13. Bouveyron C, Jacques J. Adaptive linear models for regression: improving prediction when population has changed. *Pattern Recognition Letters.* 2010; 31(14):2237–47.
14. Bastani H. Predicting with proxies: Transfer learning in high dimension. *Management Science.* 2021; 67(5):2964–84.
15. Li S, Cai TT, Li H. Transfer learning for high-dimensional linear regression: Prediction, estimation and minimax optimality. *Journal of the Royal Statistical Society Series B: Statistical Methodology.* 2022; 84(1):149–73. <https://doi.org/10.1111/rssb.12479> PMID: 35210933
16. Hsu D, Sabato S. Loss minimization and parameter estimation with heavy tails. *Journal of Machine Learning Research.* 2016; 17(18):1–40.
17. Marandi A, Ben-Tal A, den Hertog D, Melenberg B. Extending the scope of robust quadratic optimization. *INFORMS Journal on Computing.* 2022; 34(1):211–26.
18. Lange KL, Little RJ, Taylor JM. Robust statistical modeling using the t distribution. *Journal of the American Statistical Association.* 1989; 84(408):881–96.
19. Pinheiro JC, Liu C, Wu YN. Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution. *Journal of Computational and Graphical Statistics.* 2001; 10(2):249–76.
20. Yao W, Wei Y, Yu C. Robust mixture regression using the t-distribution. *Computational Statistics & Data Analysis.* 2014; 71:116–27.
21. Ubaidillah A, Notodiputro K, Kurnia A, Fitrianto A, Mangku I, editors. A robustness study of student-t distributions in regression models with application to infant birth weight data in Indonesia. *IOP Conference Series: Earth and Environmental Science;* 2017: IOP Publishing.
22. Lu K-P, Chang S-T. Robust algorithms for change-point regressions using the t-distribution. *Mathematics.* 2021; 9(19):2394.
23. Liu C, Rubin DB. ML estimation of the t distribution using EM and its extensions, ECM and ECME. *Statistica Sinica.* 1995:19–39.
24. Hastie T, Tibshirani R, Friedman JH, Friedman JH. *The elements of statistical learning: data mining, inference, and prediction:* Springer; 2009.
25. Tibshirani R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology.* 1996; 58(1):267–88.
26. Peel D, McLachlan GJ. Robust mixture modelling using the t distribution. *Statistics and computing.* 2000; 10:339–48.
27. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society: series B (methodological).* 1977; 39(1):1–22.
28. Li S, Cai T, Duan R. Targeting underrepresented populations in precision medicine: A federated transfer learning approach. *The Annals of Applied Statistics.* 2023; 17(4):2970–92. <https://doi.org/10.1214/23-AOAS1747> PMID: 39314265
29. Eaton E, Desjardins M, Lane T, editors. Modeling transfer relationships between learning tasks for improved inductive transfer. *Machine Learning and Knowledge Discovery in Databases: European*

- Conference, ECML PKDD 2008, Antwerp, Belgium, September 15–19, 2008, Proceedings, Part I 19; 2008: Springer.
30. Zhang Y, Zhu Z. Transfer Learning for High-dimensional Quantile Regression via Convolution Smoothing. arXiv preprint arXiv:221200428. 2022.
 31. Xu X, Sun H, Luo J, Cheng X, Lv W, Luo W, et al. The Pathology of Primary Familial Brain Calcification: Implications for Treatment. *Neurosci Bull.* 2023; 39(4):659–74. <https://doi.org/10.1007/s12264-022-00980-0> PMID: 36469195
 32. Marinho W, de Oliveira JRM. JAM2: A New Culprit at the Pathophysiology of Primary Familial Brain Calcification. *J Mol Neurosci.* 2021; 71(9):1723–4. <https://doi.org/10.1007/s12031-021-01816-8> PMID: 33743113
 33. Schottlaender LV, Abeti R, Jaunmuktane Z, Macmillan C, Chelban V, O'Callaghan B, et al. Bi-allelic JAM2 Variants Lead to Early-Onset Recessive Primary Familial Brain Calcification. *Am J Hum Genet.* 2020; 106(3):412–21. <https://doi.org/10.1016/j.ajhg.2020.02.007> PMID: 32142645
 34. Bazzoni G. The JAM family of junctional adhesion molecules. *Curr Opin Cell Biol.* 2003; 15(5):525–30. [https://doi.org/10.1016/s0955-0674\(03\)00104-2](https://doi.org/10.1016/s0955-0674(03)00104-2) PMID: 14519386
 35. Lounici K, Pontil M, Tsybakov AB, Van De Geer S. Taking advantage of sparsity in multi-task learning. arXiv preprint arXiv:09031468. 2009.