

## RESEARCH ARTICLE

# PCB: A pseudotemporal causality-based Bayesian approach to identify EMT-associated regulatory relationships of AS events and RBPs during breast cancer progression

Liangjie Sun<sup>1</sup>, Yushan Qiu<sup>2\*</sup>, Wai-Ki Ching<sup>1</sup>, Pu Zhao<sup>3</sup>, Quan Zou<sup>4</sup>

**1** Department of Mathematics, The University of Hong Kong, Hong Kong, China, **2** College of Mathematics and Statistics, Shenzhen University, Shenzhen, China, **3** College of Life and Health Sciences, Northeastern University, Shenyang, China, **4** Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu, China

\* [yushanqiu2526374@163.com](mailto:yushanqiu2526374@163.com)



## OPEN ACCESS

**Citation:** Sun L, Qiu Y, Ching W-K, Zhao P, Zou Q (2023) PCB: A pseudotemporal causality-based Bayesian approach to identify EMT-associated regulatory relationships of AS events and RBPs during breast cancer progression. *PLoS Comput Biol* 19(3): e1010939. <https://doi.org/10.1371/journal.pcbi.1010939>

**Editor:** Ananda Mondal, Florida International University, UNITED STATES

**Received:** July 19, 2022

**Accepted:** February 9, 2023

**Published:** March 17, 2023

**Copyright:** © 2023 Sun 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:** Processed TCGA BRCA Level 3 RNA-SeqV2 gene expression data were downloaded from the Genomic Data Commons (GDC) Legacy Archive (<https://portal.gdc.cancer.gov/legacy-archive>) The data and code for PCB are available at: [http://hkumath.hku.hk/~wkc/PCB\(data+code\).zip](http://hkumath.hku.hk/~wkc/PCB(data+code).zip).

**Funding:** YQ was supported by grants from the National Natural Science Foundation of China [grant number 62002234] the Guangdong Basic

## Abstract

During breast cancer metastasis, the developmental process epithelial-mesenchymal (EM) transition is abnormally activated. Transcriptional regulatory networks controlling EM transition are well-studied; however, alternative RNA splicing also plays a critical regulatory role during this process. Alternative splicing was proved to control the EM transition process, and RNA-binding proteins were determined to regulate alternative splicing. A comprehensive understanding of alternative splicing and the RNA-binding proteins that regulate it during EM transition and their dynamic impact on breast cancer remains largely unknown. To accurately study the dynamic regulatory relationships, time-series data of the EM transition process are essential. However, only cross-sectional data of epithelial and mesenchymal specimens are available. Therefore, we developed a pseudotemporal causality-based Bayesian (PCB) approach to infer the dynamic regulatory relationships between alternative splicing events and RNA-binding proteins. Our study sheds light on facilitating the regulatory network-based approach to identify key RNA-binding proteins or target alternative splicing events for the diagnosis or treatment of cancers. The data and code for PCB are available at: [http://hkumath.hku.hk/~wkc/PCB\(data+code\).zip](http://hkumath.hku.hk/~wkc/PCB(data+code).zip).

## Author summary

In this paper, pseudotime causality-based Bayesian (PCB) model is proposed to detect the regulatory relationships between alternative splicing events and RNA-binding proteins in the EM transition process of breast cancer. It is the first time to reveal the global role of alternative splicing events and RNA-binding proteins during EM transition, in which RNA-binding proteins can dynamically regulate alternative splicing. Furthermore, to address the challenges that lack of temporal information in sample-based transcriptomic data, we propose to decode the latent temporal information underlying cancer progression through ordering patient sample based on transcriptomic similarity, and design a

and Applied Basic Research Foundation [grant number 2019A1515111180]. WC was supported by grants from the HKRGC GRF [grant number 17301519]. PZ was supported by grants from the Fundamental Research Funds for Central Public Welfare Research Institutes of China [grant number N2120003], Special Projects of the Central Government in Guidance of Local Science and Technology Development [grant number 2022JH6/100100025]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

latent-temporal progression-based Bayesian method to infer alternative splicing regulatory network from sample based transcriptomic data of cancer patients. We inferred dynamic regulatory relationships between EM transition-associated alternative splicing events and RNA-binding proteins. Overall, we determined that ZCCHC24, PIWIL4, and TDRD10 may be key RNA-binding proteins in the EM transition process of breast cancer. In addition, we studied the dynamic regulatory relationships between alternative splicing events of the CD44 gene and RNA-binding proteins (hnRNPM, ESRP1 and ESRP2). Our study suggests that inferring dynamic progression trajectory from static expression data of tumor samples helps to uncover regulatory mechanism underlying cancer progression. In addition, our study proposes a new and effective way to characterize cancer progression by clinical transcriptomic data modeling and facilitate the incorporation of a regulatory network-based approach into precision medicine.

## Introduction

Tumor metastasis, which involves the spread of cancer cells from the original site to other organs, has always been the major cause of death in cancer patients. During tumor metastasis, tumor cells hijack the development of epithelial-mesenchymal (EM) transition to attack surrounding tissues and migrate to distant organs [1, 2]. In recent years, increasing evidences have shown that EM transition is abnormally activated in cancer cells [1, 2]. Furthermore, it is generally believed that EM transition may contribute to breast cancer metastasis. EM transition is a cellular process in which epithelial cells obtain mesenchymal phenotype and behavior after epithelial characteristics are down-regulated. EM transition often appears in biological processes such as tissue repair and wound healing. EM transition is regulated by four major regulatory systems: alternative splicing, transcriptional control, post-translational control, and noncoding RNA [3]. Here, the transcriptional regulatory network controlling EM transition has been well studied [2, 4, 5]. However, alternative RNA splicing also plays a critical regulatory role during this process. A comprehensive understanding of alternative splicing events and the RNA-binding proteins that dynamically regulate it during EM transition and their impact on breast cancer remains largely unknown.

Alternative splicing, also referred to as selective RNA splicing or differential splicing, is the process of choosing diverse splicing site combinations in a precursor messenger RNA (pre-mRNA) to produce variably spliced mRNAs. Proteins with different sequences and activities are encoded through these mRNAs, but all these mRNAs come from a single gene. In other words, alternative splicing is an important biological process by which cells can express several variants, also known as isoforms, of a single gene. Each splicing variant gives rise to a different protein with a unique structure that can perform different functions and respond to internal and environmental needs. Alternative splicing is both a vital mechanism for development and cell type-specific control of gene expression, and a mechanism for enriching proteome diversity. Numerous studies have confirmed that alternative splicing changes dynamically during EM transition, and alternative splicing regulation is crucial in EM transition [6–10]. It was reported that almost all human genes undergo alternative splicing, and alternatively spliced isoforms play different functional roles in cells [11, 12].

Cis-elements located near the pre-mRNAs variable exons regulate alternative splicing. These various cis-elements are identified by cognate RNA-binding proteins, which affect the recognition of splice sites by spliceosomes. Splicing-regulatory RNA-binding protein has a positive or negative impact on the inclusion of alternative exons. RNA-binding proteins often

interact in complexes to regulate splicing regulation [13–15]. Hence, the relationships between RNA-binding proteins and their target alternative splicing events are largely context-dependent, and cracking the “splicing code” is an urgent problem in this field [16]. It is necessary to develop a computational method to get how and to what extent diverse groups of RNA-binding proteins regulate given alternative splicing events.

Recently some effective models [17, 18] have been proposed to infer the relationships between alternative splicing events and RNA-binding proteins. However, they fail to reveal the dynamic relationships between alternative splicing events and RNA-binding proteins during the EM transition process, which is of great importance. To explore the dynamic regulatory relationships, people usually apply time-course expression data. On the other hand, the lack of sufficient temporal information is prevalent in most of the available omics data from cross-sectional studies of cancer patients, which is the main reason why it is difficult to infer the association between alternative splicing events and RNA-binding proteins.

In order to infer gene dynamics and thus infer the sequence of cellular programs, collective process dynamics can be reconstructed by reordering cells according to some measure of expression similarity, which is known as pseudotemporal ordering. For example, in [19], Wanderlust was proposed to identify cellular trajectories, which can capture nonlinear behavior. It directly constructs a collection of  $k$ -nearest neighbor graphs in high-dimensional space, and then obtains the shortest path through this collection. However, this method has certain limitations. First, Wanderlust is designed to process the data from the expression of protein markers. In this case, the number of markers is relatively small (dozens, not hundreds), and markers are selected manually according to their prior knowledge of the process involved. Secondly, when constructing cellular trajectories, Wanderlust always assumes that these processes are non-branching. Moreover, in [20], latent-temporal progression-based Bayesian (PROB) method was proposed to infer gene regulatory network, which only uses gene expression and pathological information (stage information) to infer latent-temporal progression. However, this method only considers a single progression trajectory, which is insufficient in inferring the multi branch trajectory of disease progression or cell differentiation. In addition, this method cannot handle missing data in clinical transcriptome data.

Motivated by the above problems and challenges, in this study, based on the concept of pseudotime [19] and latent temporal [20], we propose a pseudotime causality-based Bayesian (PCB) model to identify the EM transition-associated alternative splicing events and RNA-binding proteins regulatory relationships, using cross-sectional data of epithelial and mesenchymal specimens in breast cancer. Apply this model to real dataset from [21]. Using this dataset, we first found EM transition-associated alternative splicing events and RNA-binding proteins in breast cancer, and studied the dynamic regulatory relationships between these alternative splicing events and RNA-binding proteins; then we revealed the dynamic regulatory relationships between the alternative splicing events of the CD44 gene and RNA-binding proteins (heterogeneous nuclear ribonucleoprotein M, hnRNPM; epithelial splicing regulatory protein 1, ESRP1; and epithelial splicing regulatory protein 2, ESRP2) during the EM transition process.

The main contribution can be summarized as follows: 1) this paper integrates alternative splicing events and RNA-binding proteins into a dynamical system for the first time, which more clearly reveals the regulatory relationships between alternative splicing events and RNA-binding proteins; 2) our study suggests a novel and effective way to identify key regulatory relationships of alternative splicing events and RNA-binding proteins of cancer progression based on cross-sectional data; 3) our study also proposes a novel strategy to identify target alternative splicing events or key RNA-binding proteins for the diagnosis or treatment of cancers; 4) pseudotime analysis is first applied to cross-sectional alternative splicing and

RNA-binding protein expression data of epithelial and mesenchymal specimens. Although our pseudotime analysis is mainly based on the PROB model in [20], cross-sectional data with missing values is processed, and alternative splicing event and RNA-binding protein expression data are mapped into the same pseudotime trajectory.

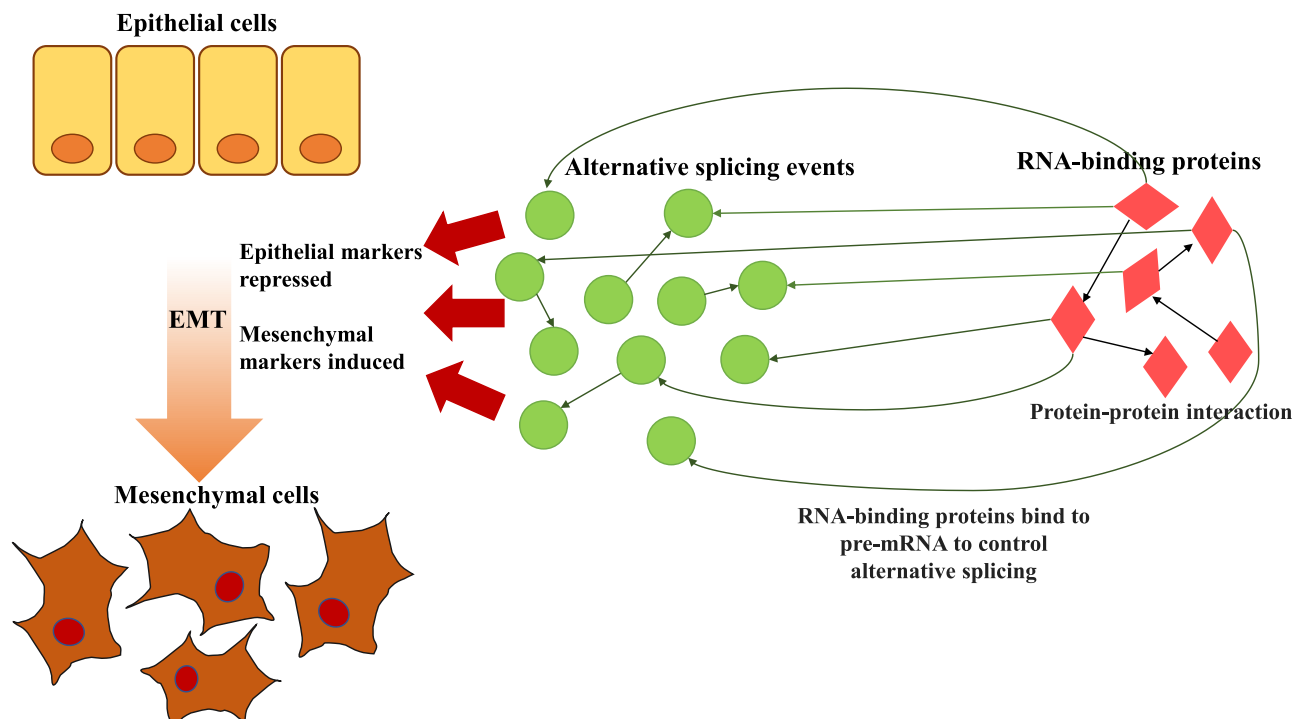
## Methods

### Problem definition and method outline

Based on the premise that alternative splicing plays an important regulatory role in the EM transition process and that RNA-binding proteins bind to pre-mRNA to regulate alternative splicing, it is necessary to propose a method that identifies the dynamic regulatory relationships between alternative splicing events and RNA-binding proteins in the EM transition process (Fig 1). To study the regulatory relationships, it is ideal to experimentally obtain EM transition time-series data. Unfortunately, collecting time-series experimental data is expensive. Currently, only cross-sectional data of epithelial and mesenchymal specimens are available.

Inspired by the above discussion, we proposed a PCB model in this paper. Based on this model, we studied the following two issues. First, we found EM transition-associated alternative splicing events and RNA-binding proteins in breast cancer and studied the dynamic regulatory relationships between these alternative splicing events and RNA-binding proteins. Second, the dynamic regulatory relationships between alternative splicing events of the CD44 gene and the RNA-binding proteins (hnRNPM, ESRP1 and ESRP2) during the EM transition process was revealed.

PCB consists of two major components. First, the time-series data requirement is solved by performing pseudotime analysis and converting static data (alternative splicing event



**Fig 1. Pictorial representation of the EM transition process by incorporating alternative splicing events and RNA-binding proteins.**

<https://doi.org/10.1371/journal.pcbi.1010939.g001>

expression and RNA-binding protein expression) into pseudotime progression data. A hypothetical example of a gene with three splice variants is given in [22] to illustrate the alternative splicing event expression. Here, variant 1 is formed by all three exons, whereas variant 2 skips the second exon and variant 3 skips the third exon. Usually, multiple variants are expressed simultaneously at any given time. Suppose that variant 1 makes up for 60% of the overall expression of the gene, variant 2 for 30% and variant 3 for 10%. Hence, there are 3 alternative splicing events of this gene, which are recorded as A1, A2 and A3 for convenience, and the expression of A1 is the overall expression of the gene multiplied by 60%, the expression of A2 is the overall expression of the gene multiplied by 30%, and the expression of A3 is the overall expression of the gene multiplied by 10%. Meanwhile, RNA-binding protein expression refers to the gene expressions of this RNA-binding protein. Then, to determine the dynamic regulatory relationships between alternative splicing events and RNA-binding proteins, we assume that the dynamic regulatory relationships can be modeled as a dynamical system. Based on our experience, we believe that the above regulatory network is always sparsely connected. According to the prior hypothesis of sparsity, the Bayesian Lasso method is used to estimate the parameters in the dynamical system. In particular, to select EM transition-associated alternative splicing events and RNA-binding proteins, a trend analysis method is applied. Specifically, based on the obtained pseudotime progression data, for each alternative splicing event (RNA-binding protein), the linear trend divided by the detrended standard deviation is calculated, and then the absolute value of this value is taken. This absolute value is defined as the score for each alternative splicing event (RNA-binding protein). The top 50 alternative splicing events with the highest scores are considered as EM transition-associated alternative splicing events, and the top 10 RNA-binding proteins with the highest scores are considered as EM transition-associated RNA-binding proteins.

## Data preprocessing

The raw data adopted in this paper is the Processed TCGA BRCA Level 3 RNA-SeqV2 gene expression data were downloaded from the Genomic Data Commons (GDC) Legacy Archive (<https://portal.gdc.cancer.gov/legacy-archive>). The data and code for PCB are available at: [http://hkumath.hku.hk/~wkc/PCB\(data+code\).zip](http://hkumath.hku.hk/~wkc/PCB(data+code).zip).

Epithelial and mesenchymal tumors were then expected to be distinguished from the original TCGA-BRCA samples. The EM transition score was calculated as the gene expression difference between E-cadherin (CDH1) and Vimentin (VIM), where CDH1 and VIM are correspondingly highly expressed in epithelial and mesenchymal cell states. Based on EM transition scores, samples with EM transition scores one standard deviation above the mean were classified as mesenchymal, and samples with EM transition scores one standard deviation below the mean were classified as epithelial. This classification identified 143 epithelial samples and 157 mesenchymal samples out of a total of 1215 TCGA-BRCA samples.

Known alternative splicing events could be classified using an annotated set of splicing events provided by the splicing analysis tool mixture-of-isoforms (MISO) model downloaded from (<http://hollywood.mit.edu.eproxy.lib.hku.hk/burgelab/miso/>) [23]. Based on the above analysis, a comprehensive list of 15753 annotated cassette exons events was created for the 300 samples (143 epithelial samples and 157 mesenchymal samples). Meanwhile, we also collected the gene expressions of 1532 RNA-binding proteins in a recent census of human RNA-binding proteins for all breast cancer (BRCA)-associated samples [24, 25].

Now, 143 epithelial specimens and 157 mesenchymal specimens were extracted from the original large dataset, and each sample contained the expressions of 15753 alternative splicing

events and the gene expressions of 1532 RNA-binding proteins ( $15753 \times 300$  data matrix and  $1532 \times 300$  data matrix). For more details, see for instance, [17, 21].

For  $15753 \times 300$  data matrix, there are many missing values. In order to fill in the missing data more accurately, we first deleted the row where the number of missing data is greater than or equal to 100. Then knnimp was used to impute the missing values. For the filled data, we calculated the row variance and deleted the row whose row variance is 0. Finally, for alternative splicing events,  $10049 \times 300$  data matrix was obtained, see [S1 Table](#) for details. Similarly, for RNA-binding proteins,  $1525 \times 300$  data matrix was obtained, see [S2 Table](#) for details. Moreover, the pathological grading (stage I-IV) and TNM staging of the 300 specimens from TCGA were given in [S3 Table](#).

### Pseudotime analysis

We applied the cross-sectional dataset of breast cancer patients in [S2 Table](#), and used the pseudotime analysis to transform the static dataset into a time-series dataset.

Specifically, we performed the pseudotime progression inference to order 300 specimens based on the whole RNA-binding protein expression profile in [S2 Table](#). Here, the epithelial and mesenchymal specimens were labeled 1 and 2, respectively. The algorithm for pseudotime analysis is described in [S1 Text](#). Using this algorithm, for each specimen we calculated its pseudotime score. According to the pseudotime score, the corresponding specimens were arranged in ascending order, and the sorted samples were mapped to a smoothed temporal trajectory. Based on the sorted TCGA barcode, we could directly sort the alternative splicing event expression profile in [S1 Table](#), and similarly mapped the sorted samples to a smoothed temporal trajectory.

### EM transition-associated alternative splicing event and RNA-binding protein selection

To obtain alternative splicing events and RNA-binding proteins associated with the EM transition process in breast cancer, we used a trend analysis technique. Based on the above pseudotime analysis, we define  $X_i(s)$  as the smoothed expression of the alternative splicing event (or RNA-binding protein) index  $i$  at the inferred pseudotime progression status  $s$ . Each index  $i$  corresponds to an alternative splicing event (RNA-binding protein). For example,  $i = 1$  of RNA-binding protein represents the RNA-binding protein A1CF. Since there are 10049 alternative splicing events and 1525 RNA-binding proteins, the index  $1 \leq i \leq 10049$  ( $1 \leq i \leq 1525$ ) of alternative splicing event (RNA-binding protein). The trend analysis method is as follows:

**Algorithm 1** The trend analysis method

- 1: Input: smoothed expression data  $X_i$ .
- 2: Fit  $X_i$  to a linear function  $L_i \times s + C_i$ . Here, the estimated coefficient is denoted as  $L_i$  and the constant is denoted as  $C_i$ .
- 3: Calculate the standard deviation of the detrended expression data, and denote as  $V_i$ . The detrended expression data  $H_i(s)$  is obtained as  $H_i(s) = X_i(s) - (L_i \times s + C_i)$ , which means that removes the best straight-line fit linear trend from  $X_i$ . Therefore,  $V_i = \text{std}(H_i(s))$ .
- 4: A score is defined,  $R_i = |L_i/V_i|$ .
- 5: Output: score  $R_i$ .

Linear trend represents a systematic increase or decrease in data over time. By removing the linear trend from the data, the fluctuation of the detrended data was considered, and such fluctuation is not caused by time change. Therefore, we believe that the higher the score  $R_i$ , the

more likely the expression of the corresponding alternative splicing event  $i$  (RNA-binding protein  $i$ ) changes significantly in the EM transition process, in other words, this alternative splicing event  $i$  (RNA-binding protein  $i$ ) is more likely to be EM transition-related. An example is given in [S2 Text](#) to clarify this method.

For each alternative splicing event  $1 \leq i \leq 10049$  (RNA-binding protein  $1 \leq i \leq 1525$ ), we can get a score according to the above algorithm. We selected alternative splicing events and RNA-binding proteins with high scores.

### Identification of the regulatory relationships between alternative splicing events and RNA-binding proteins

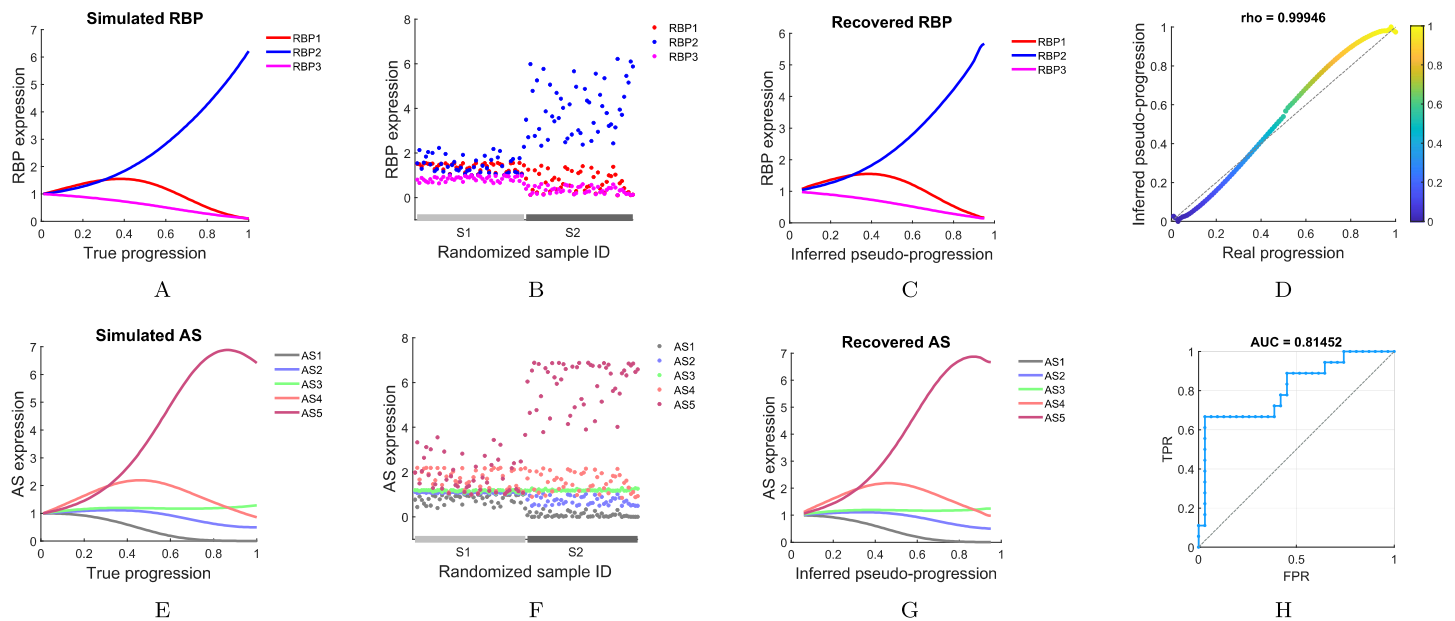
As described in [13], sequence-specific RNA-binding proteins bind to pre-mRNA to control alternative splicing. And each alternative splicing event is controlled by multiple RNA-binding proteins. Based on the mass action kinetics [20, 26, 27] and the above priori knowledge, the regulatory relationships between alternative splicing events and RNA-binding proteins can be described by means of the following dynamical system,

$$\frac{dX_i(s)}{ds} = \sum_{j \neq i} a_{ij} X_j(s) \cdot X_j(s) + \sum_{l=1}^M b_{il} X_i(s) \cdot U_l(s) - d_i X_i(s), \tag{1}$$

$$\frac{dU_l(s)}{ds} = \sum_{k \neq l} c_{lk} U_l(s) \cdot U_k(s) - d'_l U_l(s). \tag{2}$$

Here  $X_i(s)$  and  $U_l(s)$  represent the expression level of alternative splicing event  $i$ ,  $i = 1, 2, \dots, N$  and RNA-binding protein  $l$ ,  $l = 1, 2, \dots, M$  in breast cancer with pseudotime progression status  $s$ , respectively. Moreover,  $a_{ij}$  is the dynamic regulatory coefficient from alternative splicing event  $j$  to alternative splicing event  $i$ , where  $i \neq j$  and  $i, j = 1, 2, \dots, N$ ;  $b_{il}$  is the dynamic regulatory coefficient from RNA-binding protein  $l$  to alternative splicing event  $i$ , where  $i = 1, 2, \dots, N$  and  $l = 1, 2, \dots, M$ ;  $c_{lk}$  is the dynamic regulatory coefficient from RNA-binding protein  $k$  to RNA-binding protein  $l$ , where  $l \neq k$  and  $l, k = 1, 2, \dots, M$ , and  $d_i$  is the self-degradation rate of alternative splicing event  $i$ ;  $d'_l$  is the self-degradation rate of RNA-binding protein  $l$ . Since alternative splicing events rely on RNA-binding proteins, but not the other way around, [Eq \(1\)](#) has a middle term and [Eq \(2\)](#) has no middle term. For more details on this dynamical system, see [S3 Text](#). Notably, the dynamical system can be reasonably assumed to be sparsely connected.

Then, the Bayesian Lasso method is applied to estimate parameters of the above model. Although the Bayesian method is typically employed with a Gaussian prior, which allows for simple analysis and processing due to conjugation, the sparsity cannot be guaranteed using Gaussian prior. As mentioned in [28], when sparsity is taken as a key prior assumption, the Laplace distribution prior is often considered, and because of its log convexity, it also has a great computational advantage. In addition, in the next section, we also compare the Bayesian lasso method with the regular Bayesian method to further illustrate that the Bayesian Lasso method is more applicable here. The Markov chain Monte Carlo (MCMC) algorithm is applied to sample from the posterior. In this study, if the 95% credible interval (CI) of the parameter estimated by  $a_{ij}$  ( $b_{il}$ ) does not contain zero, it is considered that there is a directed edge from alternative splicing event  $j$  (RNA-binding protein  $l$ ) to alternative splicing event  $i$ ; otherwise, the edge is regarded as non-existent. Similarly, we can determine the directed edge from RNA-binding protein  $k$  to RNA-binding protein  $l$ . See [S3 Text](#) for details.



**Fig 2. Illustrating the function of our proposed model using a synthetic dataset.** A set of expression data for 3 RNA-binding proteins (RBPs) and 5 alternative splicing (AS) events in 100 cancer specimens was simulated. (A) and (E) Original RNA-binding protein (alternative splicing event) expression data. (B) and (F) Simulated cross-sectional RNA-binding protein (alternative splicing event) expression data. The sample IDs of the synthetic data were randomized and the EM transition process information was retained. (C) and (G) Recovered RNA-binding protein (alternative splicing event) expression dynamics according to inferred pseudotime progression trajectory. (D) Comparison of the inferred pseudotime progression with the true progression in the synthetic dataset, evaluated using Spearman's rank correlation coefficient ( $\rho$ ). (H) Accuracy of the dynamical systems inference evaluated using the AUC of ROC.

<https://doi.org/10.1371/journal.pcbi.1010939.g002>

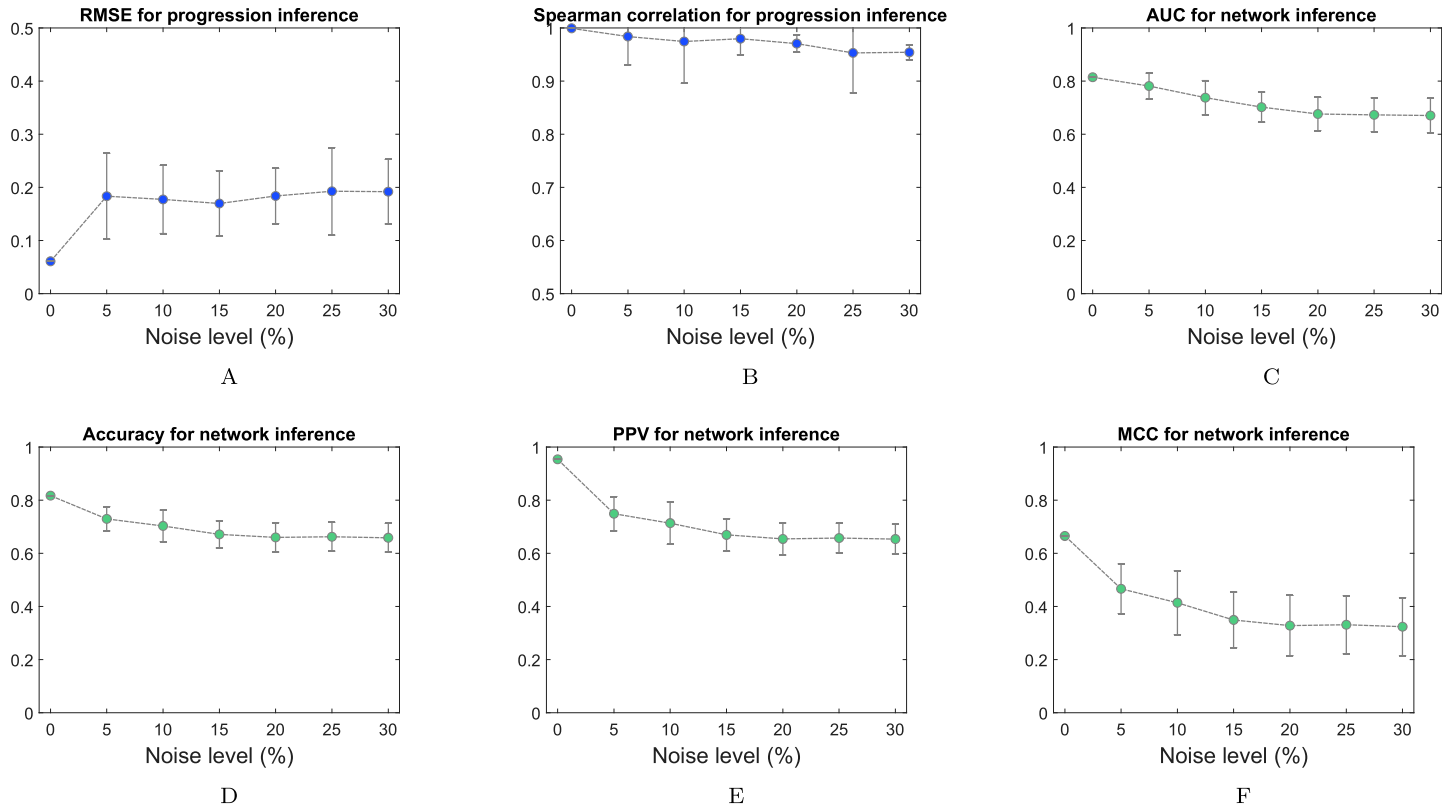
## Results and discussions

### Testing PCB with a synthetic dataset

To evaluate the effectiveness and accuracy of our proposed model PCB, a set of synthetic cross-sectional expression data was generated (S4 Text). In order to visualize the result, 3 RNA-binding proteins and 5 alternative splicing events in 100 cancer specimens were considered (Fig 2A and 2E). Based on the sample-randomized RNA-binding protein expression profile (Fig 2B), we evaluated whether the proposed method could accurately order the samples. Based on the sorted sample data, we further evaluated whether the proposed method could effectively reconstruct the regulatory relationships between alternative splicing events and RNA-binding proteins.

First, we calculated the pseudotime score of each specimen and used this score to sort the samples. Comparing the inferred pseudotime score of each specimen with the true progression (Fig 2D), the proposed model correctly recovered the real order of samples (Spearman's  $\rho = 0.99946$ ). Moreover, along with the pseudotime progression, both RNA-binding protein expression dynamics (Fig 2C) and alternative splicing event expression dynamics (Fig 2G) showed contours that are very similar to the initial data. Then, by the Bayesian Lasso method, we reconstructed the regulatory relationships between alternative splicing events and RNA-binding proteins. Compared with the real dynamical system, the area under the curve (AUC) of receiver operating characteristic (ROC) with/without interactions was used to assess the accuracy of the dynamical system inference. Here, the AUC value reached 81.452% (Fig 2H).

The robustness of the model to the measured variables in the data was illustrated by numerical verification and mathematical proof. First, multiplicative exponential noise was used to randomly perturb both alternative splicing event and RNA-binding protein expressions to



**Fig 3. Evaluation metrics were used to verify the robustness of PCB against a series of variations in the data using multiplicative exponential noise (with CVs ranging from 0% to 30%).** (A) and (B) The root mean square error (RMSE) and Spearman correlation coefficients were employed to evaluate the accuracy of the temporal progression inference. (C),(D),(E) and (F) The AUC, accuracy, positive predictive value (PPV) and Matthews correlation coefficient (MCC) were employed to evaluate the robustness of the reconstructed regulatory relationships between alternative splicing events and RNA-binding proteins.

<https://doi.org/10.1371/journal.pcbi.1010939.g003>

simulate diverse levels of measurement variabilities in the data, which resulted in a series of coefficients of variation (CVs). Here, we assumed that noise is generated from an exponential distribution with a mean  $\mu$ , where  $0\% \leq \mu \leq 10\%$ . Sample IDs were randomly assigned to simulate sample-based snapshots of alternative splicing event expression data (RNA-binding protein expression data), but still retained each specimen’s EM transition process information. PCB was applied to infer the regulatory relationships of each dataset. Evaluation metrics were used to verify the robustness of PCB against a series of changes in the data (with CVs ranging from 0% to 30%) (Fig 3). Moreover, the robustness of the model to the measured variables in the data was illustrated as follows.

**Theorem 1** Suppose there are two trajectories of pseudotime progression  $s(r)$  and  $\tilde{s}(r)$  with the same root  $r \in I = [0, 1]$ . Define  $\|\tilde{s} - s\|_{L^2} = (\int_I |\tilde{s} - s|^2 dr)^{1/2}$ . If  $(X_i(s), U_i(s), a_{ij}, b_{il}, c_{lk})$  and  $(X_i(\tilde{s}), U_i(\tilde{s}), \tilde{a}_{ij}, \tilde{b}_{il}, \tilde{c}_{lk})$  both satisfy the equations of the pseudotime progression-dependent dynamical system, i.e.,

$$\frac{dX_i(s)}{ds} = \sum_{j \neq i} a_{ij} X_j(s) \cdot X_j(s) + \sum_{l=1}^M b_{il} X_i(s) \cdot U_l(s) - d_i X_i(s), \tag{3}$$

$$\frac{dU_i(s)}{ds} = \sum_{k \neq i} c_{ik} U_i(s) \cdot U_k(s) - d'_i U_i(s), \tag{4}$$

$$\frac{dX_i(\tilde{s})}{d\tilde{s}} = \sum_{j \neq i} \tilde{a}_{ij} X_j(\tilde{s}) \cdot X_j(\tilde{s}) + \sum_{l=1}^M \tilde{b}_{il} X_l(\tilde{s}) \cdot U_l(\tilde{s}) - \tilde{d}_i X_i(\tilde{s}), \tag{5}$$

$$\frac{dU_l(\tilde{s})}{d\tilde{s}} = \sum_{k \neq l} \tilde{c}_{lk} U_k(\tilde{s}) \cdot U_k(\tilde{s}) - \tilde{d}'_l U_l(\tilde{s}), \tag{6}$$

where  $i = 1, 2, \dots, N$  and  $l = 1, 2, \dots, M$ . Then we have

$$\lim_{\|\tilde{s}-s\|_{L^2} \rightarrow 0} \left( \sum_{j=1}^N (\tilde{a}_{ij} - a_{ij})^2 + \sum_{l=1}^M (\tilde{b}_{il} - b_{il})^2 \right) = 0, \tag{7}$$

and

$$\lim_{\|\tilde{s}-s\|_{L^2} \rightarrow 0} \left( \sum_{k=1}^M (\tilde{c}_{lk} - c_{lk})^2 \right) = 0. \tag{8}$$

The details of mathematical proof can be found in [S5 Text](#).

By testing PCB with a synthetic dataset, we found that both the pseudotime progression trajectory and network structure of the dynamic regulatory relationship model inferred by PCB were quite robust to the variation in the data (Figs 2 and 3).

### Identifying EM transition-associated alternative splicing events and RNA-binding proteins during breast cancer progression

The model was applied to a dataset of breast cancer from [21]. To find alternative splicing events and RNA-binding proteins that were strongly associated with the EM transition process, we first used pseudotime analysis to obtain the ordered RNA-binding protein (alternative splicing event) expression data. The sorted specimens are shown in [S4 Table](#). Here we find that after the pseudotime analysis, all samples of epithelial state are at the beginning of the trajectory, and all samples of the mesenchymal state are at the tail of the trajectory, and most of the samples at the beginning of the trajectory have no distal metastasis (M0). However, this order seems not to be related to the pathological stage. Then using Algorithm 1, we calculated the score of each RNA-binding protein (alternative splicing event). Finally, according to the scores obtained, we selected the top 50 alternative splicing events and the top 10 RNA-binding proteins (see for more details in [S5 Table](#)).

We found that most of these 10 RNA-binding proteins increased during the EM transition, and only LARP4 decreased (see [S1 Fig](#) for details). To confirm that the RNA-binding proteins we selected were indeed related to EM transition in breast cancer, we consulted relevant literature. Among the 10 RNA-binding proteins we identified, SMAD6 and PIWIL4 have been confirmed in the literature that they were related to EM transition in breast cancer (see details in [Table 1](#)), and the roles of these two RNA-binding proteins in the EM transition process (activation or inhibition) were consistent with our reconstructed expression dynamics ([S1 Fig](#)). Additionally, ZFP36, LARP4 and ZCCHC24 have been confirmed to be associated with breast cancer metastasis and invasion (see details in [Table 1](#)). Moreover, in [29–34], it was also mentioned that the seven RNA-binding proteins (ZFP36, ZCCHC24, ZFP36L2, TDRD10, PTRF, CSDA and DZIP1) were related to the EM transition process, although these studies did not evaluate the EM transition process in breast cancer. Therefore, eight of the RNA-binding proteins we identified (ZFP36, LARP4, ZCCHC24, ZFP36L2, TDRD10, PTRF, CSDA, and

**Table 1. RNA-binding proteins (RBPs) reported in previously published literature.**

RBP	EM transition regulation	Description of RBPs role in breast cancer and relevant references
ZFP36	Activator	In vivo and in vitro, NNT-AS1 knockout can inhibit the progression and metastasis of breast cancer by interacting with ZFP36 [36].
LARP4	Inhibitor	LARP4 the inhibits migration and invasion of cancer cells [37].
SMAD6	Activator	[38] confirmed the importance of BMP signaling in the EM transition process and in vivo invasion. Overexpression of SMAD6 in the estrogen receptor breast cancer cell lines can effectively block BMP signalling.
ZCCHC24	Activator	ZCCHC24 seems to change the structure of breast cancer cells, increasing the invasiveness of the fatal disease [39].
PIWIL4	Activator	PIWIL4 is necessary for the EM transition process [40]. PIWIL4 is a novel estrogen receptor signaling regulator that can be targeted to inhibit the growth and migration of breast cancer [41].

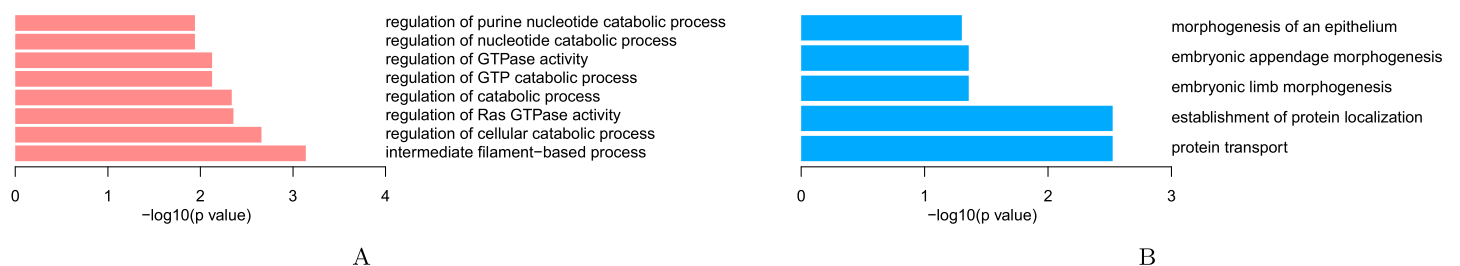
<https://doi.org/10.1371/journal.pcbi.1010939.t001>

DZIP1) could provide biologists with a new set of RNA-binding proteins to test the regulatory role of the EM transition process in breast cancer.

Then, for the top 50 alternative splicing events, a heatmap with hierarchical clustering was shown in [S2 Fig](#). Moreover, for these two groups of alternative splicing events, we analyzed the network ontology of their corresponding genes using NOA software [35] (see details in [Fig 4](#)), which showed that the alternative splicing events we found were related to EM transition [42–44]. For example, [43] mentioned that intermediate filaments plays a vital role in EM transition. Ras GTPase-activating protein SH3 domain binding protein 1 (G3BP1) is an important Ras mediator and [44] indicated that upregulation of G3BP1 promotes the EM transition process in breast cancer cells.

Moreover, we used dbEMT [45] (a literature based genetic resource for exploring EM transition related human genes) to search the corresponding genes for 50 alternative splicing events, and found that seven of them (MAP4K4, NUMB, CTNND1, NF1, FGFR1, MAP3K7, EXOC7) were clearly identified as EM transition-related genes. In addition, we found that about 23 corresponding genes were mentioned to be related to the EM transition process in the literature (See [Table 2](#) for details).

To understand the biological role and potential functions of the 50 alternative splicing events, functional enrichment analysis of the corresponding genes was conducted using Metascape (<https://metascape.org>). Pathway and process enrichment analysis has been carried out with Gene Ontology (GO) Molecular Functions. All genes in the genome have been used as the enrichment background. Terms with a  $p$ -value  $< 0.01$ , a minimum count of 3, and an enrichment factor  $> 1.5$  were collected and grouped into clusters based on their membership similarities. The results of GO analysis were displayed in [Fig 5A and 5B](#). And significant



**Fig 4. Network ontology of corresponding genes for 50 alternative splicing events.** (A) Corresponding ascending genes. (B) The corresponding descending genes.

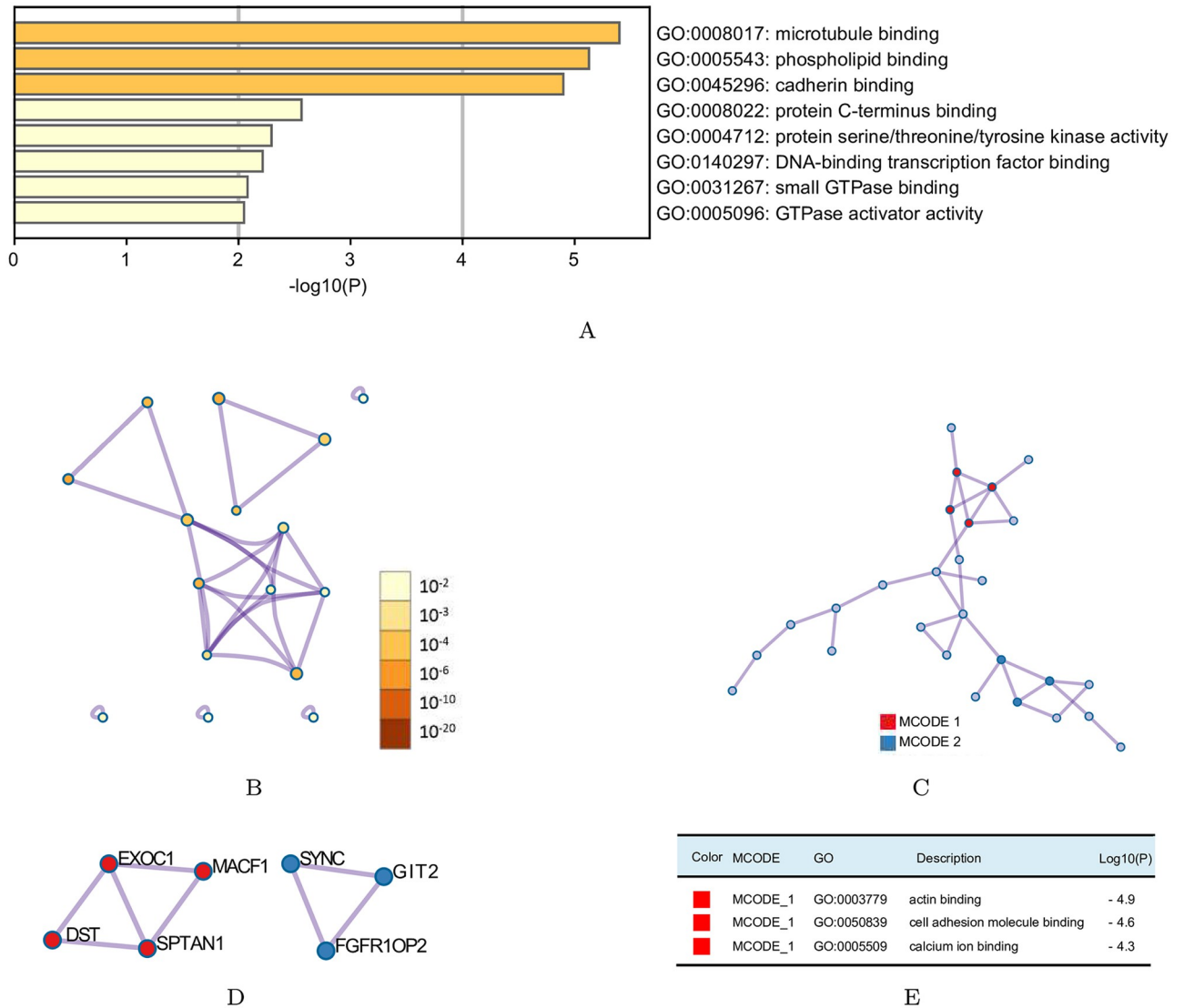
<https://doi.org/10.1371/journal.pcbi.1010939.g004>

Table 2. Genes reported in previously published literature.

Gene	EM transition regulation	Description of gene role in the EM transition (EMT) process and relevant references
MAP4K4	Activator	MAP4K4 primarily promotes the EMT process and invasiveness of hepatocellular carcinoma cells by activating JNK and NF-kappaB signaling [50].
ERP29	Inhibitor	ERP29 negatively regulates the EMT process in breast cancer cells [51].
NUMB	Inhibitor	By activating the Notch signaling, the deletion of NUMB promotes the EMT program in triple-negative breast cancer cells [52].
ARNT	Inhibitor	ARNT, as a inhibitor, prevents the progression of EMT by reducing the expression of metastasis related gene [53].
CTNND1	Inhibitor	Through E-cadherin, CTNND1 may regulate EMT in oral squamous cell carcinoma [54].
TCF12	Activator	TCF12 induces the EMT process in hepatocellular carcinoma cells [55].
KIAA1217	Activator	KIAA1217 significantly promotes cell migration and invasion by inducing the EMT process in vitro [56].
EPB41	Inhibitor	Knockout of EPB41 with shRNAs can induce morphological changes associated with the EMT process in non-small-cell lung cancer PC9 or A549 cells [57].
CLSTN1	Inhibitor	A CLSTN1 short isoform can inhibit EMT in breast cancer cells [58].
NF1	Inhibitor	Neurofibromin, encoded by NF1 gene, may inhibit the EMT process [59].
FGFR1	Activator	Activation of FGFR1 can promote the EMT process in breast cancers [60].
LRRFIP2	Activator	The WNT pathway is another signaling pathway that crosstalks the TGF- $\beta$ pathway and promotes EMT, and LRRFIP2 is an activator of canonical WNT signaling [61, 62].
SPTAN1	Activator&Inhibitor	SPTAN1 interacts with ankyrin, E-cadherin, and Na/K-ATPase. Therefore, it may affect the EMT process and metastasis [63].
PLOD2	Activator	PLOD2 promotes the EMT process, which is an upstream factor regulating the PI3K/Akt signaling pathways in glioma cells [64].
MAP3K7	Activator	Inhibition of MAP3K7 can block the EMT process in vitro [65].
MBNL1	Activator&Inhibitor	Knockout of MBNL1 induces EMT-like morphological changes in the HCT-116 cells while enhancing cell motility, upregulating Snail expression, and downregulating E-cadherin. In contrast, the ectopic overexpression of MBNL1 inhibits EMT, which is characterized by E-cadherin upregulation and Snail expression downregulation [66].
SPAG9	Activator	SPAG9 regulates the EMT process in bladder transitional carcinoma [67].
MACF1	Activator	MACF1 knockout induces the conversion of mesenchymal and epithelial biomarkers expression. Expression levels N-cadherin and TGF $\beta$ are significantly reduced, while expression levels of E-cadherin and SMAD-7 are elevated, suggesting that MACF1 has influence on the EMT process in this particular case [68].
EXOC7	Inhibitor	The expression of EXOC7 epithelial isoform can affect the levels of key EMT transcriptional regulators and is sufficient to drive the transition to epithelial phenotypes [69].
PBX1	Activator	PBX1 promotes EMT induction in lung cancer [70].
MYOF	Activator	The loss of MYOF shifts the movement of breast cancer cells towards collective migration and promotes changes in the shape of the mesenchymal-to-epithelial [71, 72].
GIT2	Inhibitor	Loss of GIT2 induces the EMT process by promoting the expression of Zeb1 [73].
NCOR2	Inhibitor	Reduce the activation of Notch signaling pathway by upregulating NCOR2 and Notch has important function on promoting the development of EMT and tumor progression [74, 75].

<https://doi.org/10.1371/journal.pcbi.1010939.t002>

modules of Protein-protein interaction enrichment analysis (PPI) were identified in Fig 5. Furthermore, the molecular complex detection (MCODE) algorithm [46] was applied to analyze clusters of the PPI networks (Fig 5D and 5E). We could see from Fig 5 that the main biological functions of the 50 alternative splicing candidate genes focused on binding with microtubule, phospholipid, cadherin, protein C-terminus, small GTPase and DNA-binding transcription factor, protein serine/threonine/tyrosine kinase activity and GTPase activator activity. PPI showed that the interaction between the 27 alternative splicing events among the 50 alternative splicing events were recruited on cell adhesion molecule binding, cadherin binding and structural molecule activity. Two molecular complexes were detected by the MCODE algorithm. Complex 1, including SPTAN1, EXOC1, MACF1 and DST, was involved in actin binding, cell adhesion molecule binding and calcium ion binding. Therefore, the above analysis revealed that the top 50 alternative splicing events mainly regulated the EM transition process by affecting cytoskeletal rearrangement and cell adhesion. Since, the changes of cytoskeletal and cell



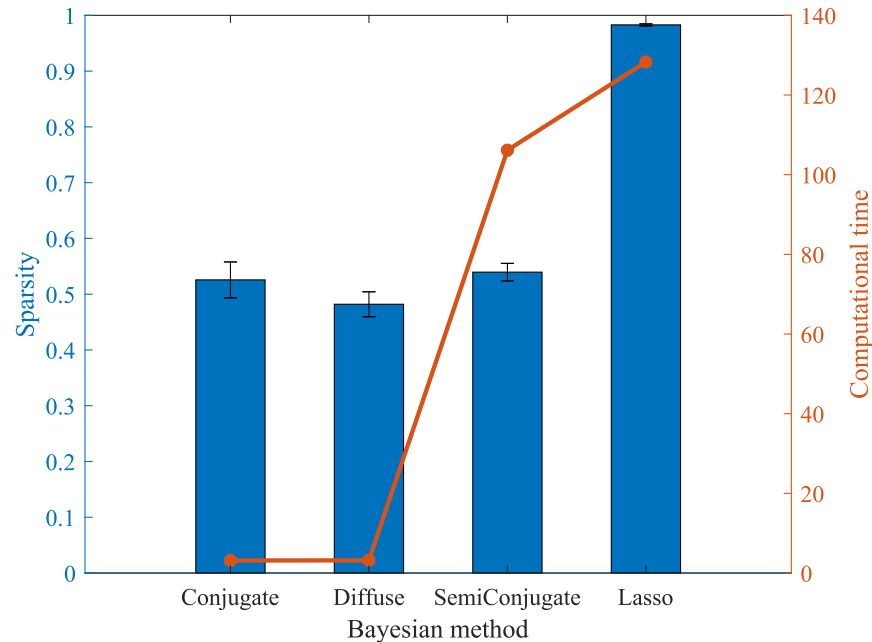
**Fig 5. Enrichment analysis of the marker genes identified by PCB (Metascape).** (A) Pathway and process enrichment analysis has been carried out with Gene Ontology (GO) Molecular Functions. Bar graph of enriched terms across the marker genes, colored by *p*-values. (B) Network of enriched terms across the marker genes: colored by *p*-value, where terms containing more genes tend to have a more significant *p*-value. (C) Protein-protein interaction network and the MCODE component identified in the marker genes. (D) Pathway and process enrichment analysis has been applied to the MCODE component. The three best-scoring terms by *p*-value have been retained as the functional description of the MCODE components, shown in Table (E).

<https://doi.org/10.1371/journal.pcbi.1010939.g005>

adhesion are the key characteristics in the EM transition process [47–49], suggesting that the 50 alternative splicing events we got is crucial for the regulation of the EM transition process, and thus also proves the reliability of our calculating method.

The above evidence shows that our method effectively selected alternative splicing events and RNA-binding proteins related to EM transition.

We then inferred the dynamic regulatory relationships of the above 50 alternative splicing events and 10 RNA-binding proteins. Here, we first compare the Bayesian Lasso method with other Bayesian methods (Bayesian Conjugate, Bayesian SemiConjugate, and Bayesian Diffuse) in estimating the parameters of the dynamical system (Eqs (1) and (2)), mainly considering



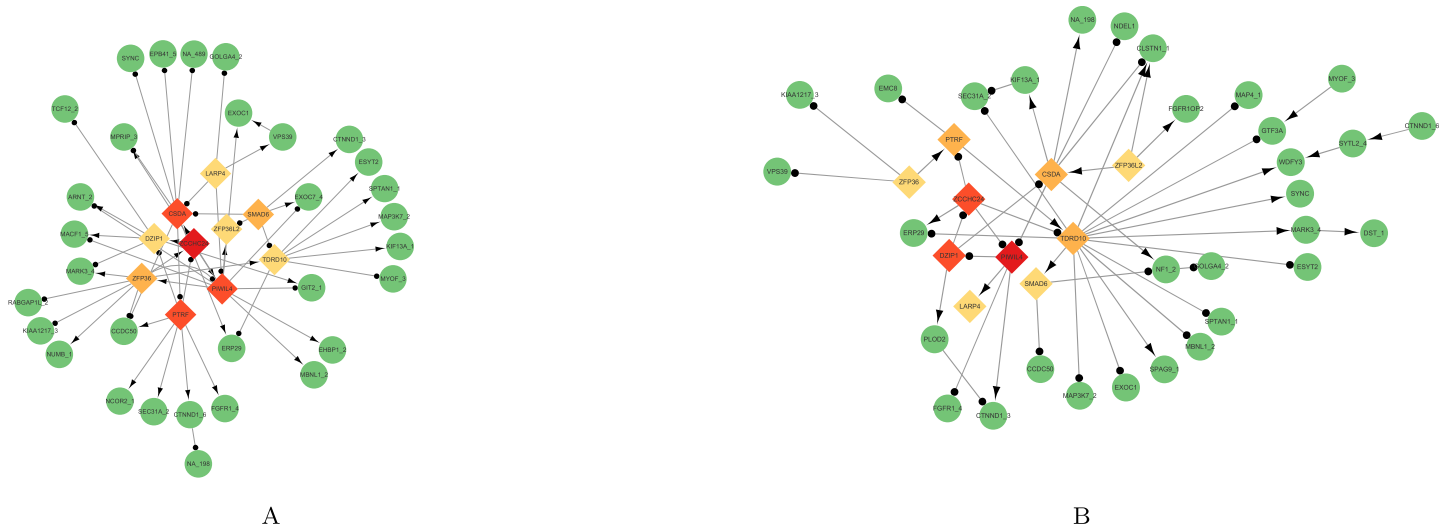
**Fig 6. Computational time and sparsity of Bayesian Lasso method and other Bayesian methods.**

<https://doi.org/10.1371/journal.pcbi.1010939.g006>

computational time and sparsity. We obtained the computational time of parameter estimates for two dynamical systems (epithelial state and mesenchymal state) and found that the computational time was not long. Specifically, Bayesian Conjugate method and Bayesian Diffuse method are about 3 seconds, Bayesian SemiConjugate is about 106 seconds, and Bayesian Lasso method is about 128 seconds. The sparsity =  $1 - |\text{interactions}| / (N^2 + NM + M^2)$ , where  $N$  is the number of alternative splicing events and  $M$  is the number of RNA-binding proteins, which refers to the degree to which parameters contain zero values (no interaction). For each Bayesian method, we still considered the 95% CI of the parameter estimated and we calculated the sparsity of the parameters estimated by the dynamical system in the epithelial state and the sparsity of the parameters estimated by the dynamical system in the mesenchymal state. According to Fig 6, using other Bayesian methods (Bayesian Conjugate, Bayesian SemiConjugate, and Bayesian Diffuse) cannot guarantee sparsity, but the dynamical system is sparsely connected according to prior knowledge.

Therefore, despite the computational time is longer, we still used Bayesian Lasso method to estimate parameters of the dynamical system (Eqs (1) and (2)).

Based on (2), we first analyzed the mutual regulatory relationships between 10 RNA-binding proteins and obtained two RNA-binding protein networks (S3 Fig). We found that the dynamic regulatory relationships between these 10 RNA-binding proteins were completely different in the epithelial state and the mesenchymal state (see S6 Table for details). Subsequently, we used the Maximal Clique Centrality (MCC) method proposed by [76] to rank the nodes in the two RNA-binding protein networks, respectively. Based on the MCC method, ZCCHC24 was considered the most influential RNA-binding protein in the RNA-binding protein network during the epithelial state (see S7 Table for details) and PIWIL4 was considered as the most influential RNA-binding protein in the RNA-binding protein network during the mesenchymal state (see S7 Table for details). Notably, [40] confirmed that MDA-MB-231 cells lacking PIWIL4 abandoned the mesenchymal cell fate and regained the key characteristics



**Fig 7. Dynamic regulatory relationship between 50 alternative splicing events and 10 RNA-binding proteins.** (A) Dynamic regulatory relationships between 50 alternative splicing events and 10 RNA-binding proteins in the epithelial state. ZCCHC24 was identified as the most influential RNA-binding protein. (B) Dynamic regulatory relationships between 50 alternative splicing events and 10 RNA-binding proteins in the mesenchymal state. PIWIL4 was identified as the most influential RNA-binding protein.

<https://doi.org/10.1371/journal.pcbi.1010939.g007>

of epithelial cell fate. Moreover, PIWIL4 is highly expressed in the cytoplasm of MDA-MB-231 cells derived from breast cancer. PIWIL4 is necessary for the transformation of epithelial cells into mesenchymal cells and the migration ability of MDA-MB-231 cells.

Based on (1), we analyzed the mutual regulatory relationships between 50 alternative splicing events and 10 RNA-binding proteins, and we obtained two alternative splicing regulatory networks, respectively. We found that an alternative splicing event could be regulated by multiple RNA-binding proteins, and when multiple RNA-binding proteins affect an alternative splicing event, these RNA-binding proteins play an antagonistic or coordinating role. For example, in the epithelial state, three RNA-binding proteins (CSDA, ZFP36 and PTRF) jointly regulate the alternative splicing event CCDC50, in which CSDA and ZFP36 inhibit the expression of CCDC50, whereas PTRF promotes the expression of CCDC50. However, one RNA-binding protein can also affect multiple alternative splicing events. For example, in the mesenchymal state, the RNA-binding protein TDRD10 is associated with multiple alternative splicing events. Finally, we obtained Fig 7 by integrating RNA-binding protein network information with alternative splicing regulatory network information. We found that PIWIL4 and TDRD10 have the largest out-degree values in the epithelial and mesenchymal states, respectively.

Therefore, the above analysis revealed that ZCCHC24, PIWIL4, and TDRD10 may be key RNA-binding proteins in EM transition of breast cancer. In particular, [40, 41] confirmed that PIWIL4 was related to EM transition in breast cancer.

### Regulation of alternative splicing events of the CD44 gene in breast cancer

The CD44 pre-mRNA contains 19 exons, nine of which are alternatively spliced. Recent studies confirmed that alternative splicing events of the CD44 gene have a causal relationship to the EM transition process and breast cancer metastasis [7, 9, 77, 78]. By studying the regulation mechanism of alternative splicing events of the CD44 gene, the antagonism between hnRNPM and ESRP1 has also been determined. Numerous studies have shown that hnRNPM promotes

CD44 variable exon skipping, which is conducive to forming the mesenchymal phenotype, whereas ESRP1 stimulates the CD44 variable exon inclusion body and promotes an epithelial cellular state [7, 9, 77, 79]. Furthermore, in [8], the RNA-binding proteins (ESRP1 and ESRP2) were identified as key regulators of splicing events related to the EM transition process. In this part, we analyzed alternative splicing events of the CD44 gene coregulated by hnRNPM, ESRP1 and ESRP2.

We still applied the model to the dataset from [21]. The pseudotime progression inference was first performed. To identify the dynamic regulatory relationships between alternative splicing events and RNA-binding proteins, we collected nine alternative splicing events of the gene CD44 and three RNA-binding proteins (hnRNPM, ESRP1 and ESRP2). We obtained the reconstructed expression dynamics of nine alternative splicing events of the CD44 gene and the reconstructed expression dynamics of hnRNPM, ESRP1 and ESRP2 (see details in Figs 8 and 9), respectively. The Wilcoxon rank-sum test (two-tailed)  $p$ -values were calculated to evaluate statistical significance. The  $p$  values  $<0.05$  were considered significant.

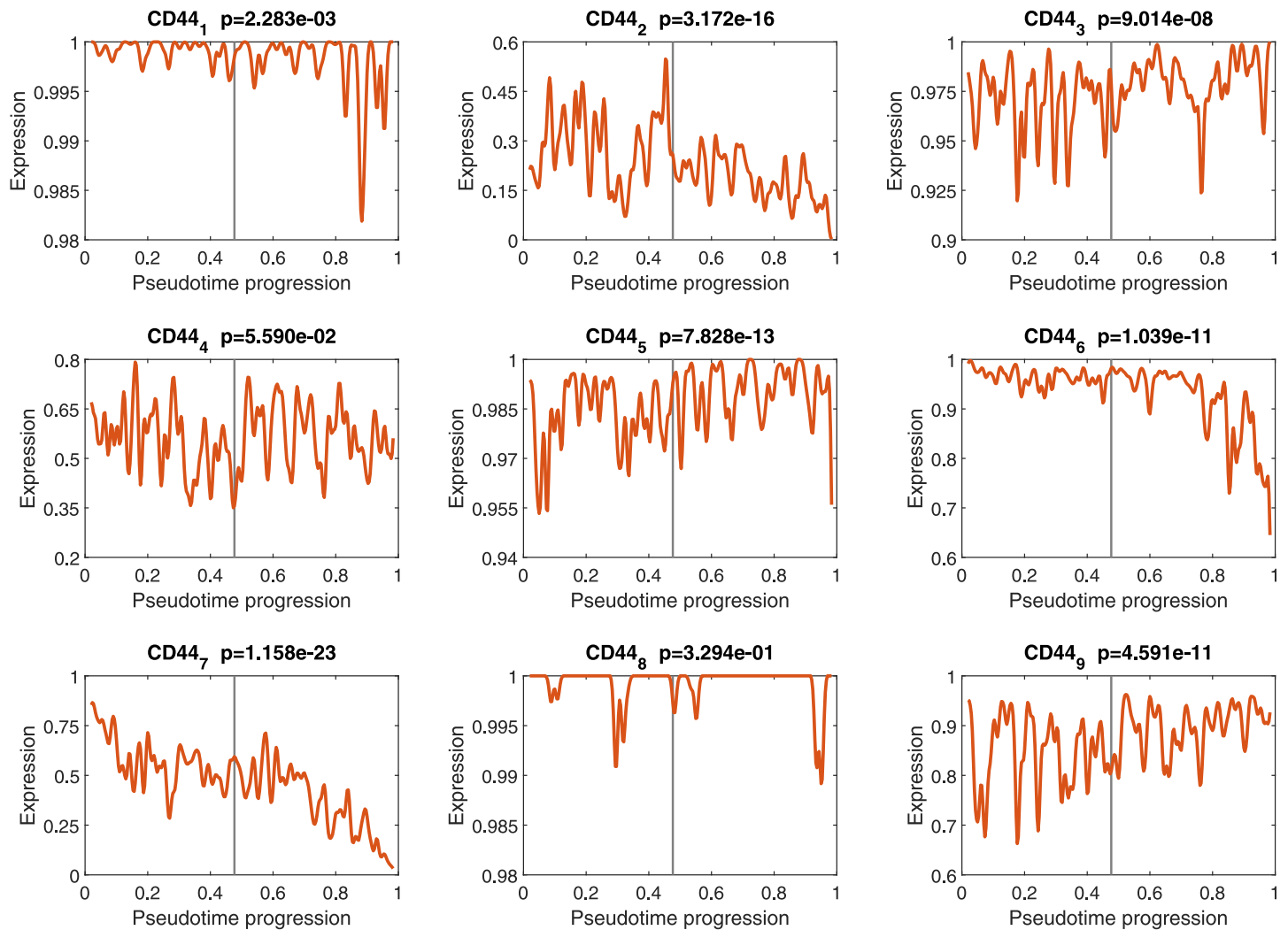
Here, we found the  $p$ -values of CD44<sub>4</sub> and CD44<sub>8</sub> were greater than 0.05. Therefore, when we reconstructed the alternative splicing regulatory networks, we only considered CD44<sub>1</sub>, CD44<sub>2</sub>, CD44<sub>3</sub>, CD44<sub>5</sub>, CD44<sub>6</sub>, CD44<sub>7</sub>, and CD44<sub>9</sub>. In the epithelial and mesenchymal samples, we obtained the corresponding reconstructed regulatory relationships between alternative splicing events and RNA-binding proteins (all adjacency matrices are given in S8 Table).

According to Fig 9 and the reconstructed regulatory relationships between alternative splicing events and RNA-binding proteins in epithelial and mesenchymal samples, we found that hnRNPM was ubiquitously expressed during the EM transition, but played a role in the state of mesenchymal cells to regulate the alternative splicing events of the CD44 gene. This finding is consistent with the result in [77]. In addition, both ESRP1 and ESRP2 were down-regulated during the EM transition process. More specifically, the expression levels of both ESRP1 and ESRP2 were high in precancerous lesions and carcinoma in situ, but their expressions were down-regulated during invasion. This phenomenon is consistent with previous experimental results [80]. These findings indicate that down-regulation of ESRP1 and ESRP2 is closely related to the motile phenotype of cancer cells.

According to Fig 8, the expression levels of the alternative splicing events CD44<sub>2</sub>, CD44<sub>6</sub> and CD44<sub>7</sub> were down-regulated. In particular, the expression levels of CD44<sub>6</sub> and CD44<sub>7</sub> were significantly down-regulated during the EM transition. However, the alternative splicing events CD44<sub>5</sub> and CD44<sub>9</sub> were up-regulated in this process. These phenomena may be related to the transformation of CD44 expression from variant subtype (CD44<sub>v</sub>) to standard subtype (CD44<sub>s</sub>) during the EM transition process [7], which is worthy of further study.

## Conclusions

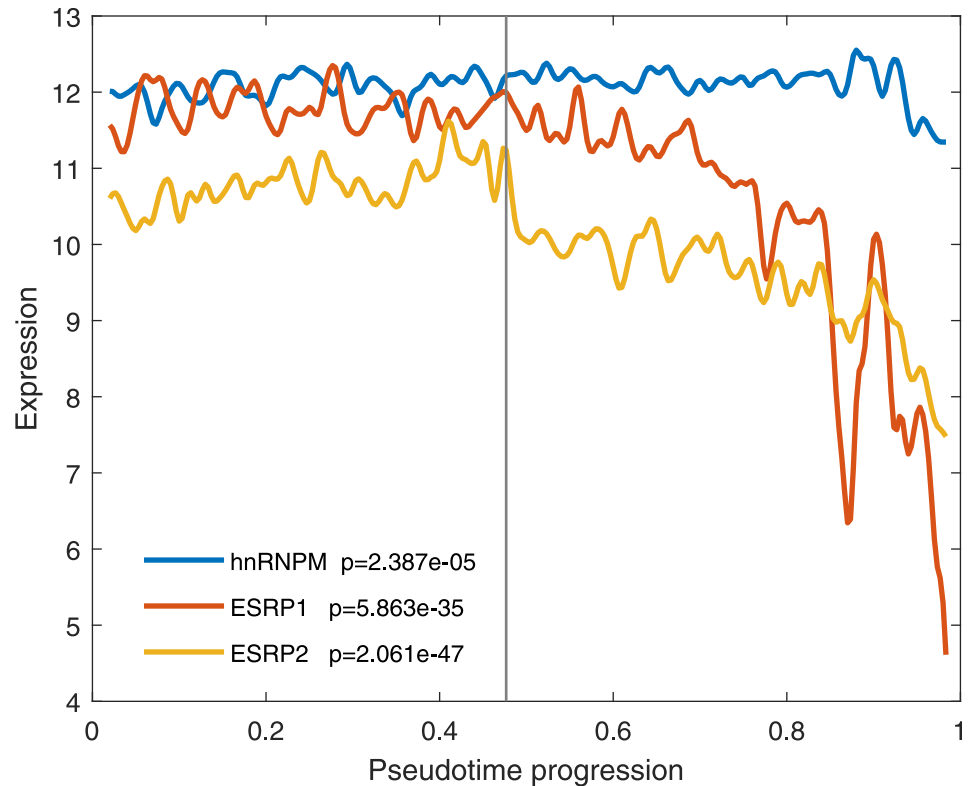
Previous studies have focused on transcriptional regulation of EM transition. However, the role of global dynamic alternative splicing regulation in EM transition remains relatively uncharacterized. It is well known that a comprehensive understanding of alternative splicing and its regulation is essential for understanding its biological impact on the disease process. In this study, we attempted to characterize the alternative splicing differences and its dynamic regulation between epithelial and mesenchymal breast tumors from TCGA. To study these regulatory relationships, a PCB model was proposed. First, the time-series data of alternative splicing expression and RNA-binding protein expression were obtained by pseudotime analysis. Then, a trend analysis technique was used to select EM transition-associated alternative splicing events and RNA-binding proteins. According to the selected alternative splicing events and RNA-binding proteins, we used a dynamical system to simulate the dynamic



**Fig 8. Reconstructed expression dynamics of alternative splicing events of the CD44 gene.**

<https://doi.org/10.1371/journal.pcbi.1010939.g008>

regulatory relationships between alternative splicing events and RNA-binding proteins, and used the Bayesian Lasso method to estimate the parameters in the dynamical system. Finally, we analyzed the dynamic regulatory relationships between alternative splicing events and RNA-binding proteins, and identified the key RNA-binding proteins related to EM transition. PCB is a novel and effective tool based on dynamical system representation of alternative splicing events and RNA-binding proteins interactions during cancer progression. PCB can be used to generate experimentally testable hypotheses and to identify network-based alternative splicing biomarkers for predicting cancer prognosis and treatment response. Besides cross-sectional bulk transcriptomic data, our method could also be used to time-course scRNA-seq data. Clinical transcriptomic data of cancer patients provide an alternative way to infer alternative splicing regulation networks underlying cancer progression. Our study also sheds light on facilitating the regulatory network-based approach to identify target alternative splicing events or key RNA-binding proteins for the diagnosis or treatment of cancers.



**Fig 9. Reconstructed expression dynamics of hnRNPM, ESRP1 and ESRP2.**

<https://doi.org/10.1371/journal.pcbi.1010939.g009>

## Supporting information

**S1 Fig. Reconstructed expression dynamics of 10 selected RNA-binding proteins.** Most of these 10 RNA-binding proteins increased during the EM transition, and only LARP4 decreased.

(PDF)

**S2 Fig. A heatmap with hierarchical clustering of 50 alternative splicing events.** 50 alternative splicing events were clearly clustered into two groups: a descending group and an ascending group.

(PDF)

**S3 Fig. The mutual regulatory relationships between 10 RNA-binding proteins.** The MCC value was calculated for each RNA-binding protein. The larger the MCC value, the darker the color in the figure. (A) RNA-binding protein network in the epithelial state. Here, ZCCHC24 was identified as the most influential RNA-binding protein. (B) RNA-binding protein network in the mesenchymal state. Here, PIWIL4 was identified as the most influential RNA-binding protein.

(PDF)

**S1 Table. 10049 × 300 data matrix, 10049 annotated alternative splicing events for the 300 samples (143 epithelial samples and 157 mesenchymal samples).**

(XLSX)

**S2 Table. 1525 × 300 data matrix, 1525 RNA-binding proteins for the 300 samples (143 epithelial samples and 157 mesenchymal samples).**

(XLSX)

**S3 Table. 300 specimens' pathological classification (Stage I-IV) and TNM stage.**

(XLSX)

**S4 Table. The ordered specimens.**

(XLSX)

**S5 Table. The 50 alternative splicing events and 10 RNA-binding proteins selected in this study.**

(XLSX)

**S6 Table. The dynamic regulatory relationships between the 10 RNA-binding proteins in the epithelial state and the mesenchymal state, respectively.**

(XLSX)

**S7 Table. The MCC of the 10 RNA-binding proteins in the epithelial state and the mesenchymal state, respectively. ZCCHC24 was identified as the most influential RNA-binding protein in the RNA-binding protein network during the epithelial state and PIWIL4 was identified as the most influential RNA-binding protein in the RNA-binding protein network during the mesenchymal state.**

(XLSX)

**S8 Table. The reconstructed regulatory relationships between alternative splicing events of the CD44 gene and RNA-binding proteins (hnRNPM, ESRP1 and ESRP2).**

(XLSX)

**S1 Text. The algorithm for pseudotime analysis.**

(PDF)

**S2 Text. An example for trend analysis.**

(PDF)

**S3 Text. The model of the regulatory relationships between alternative splicing events and RNA-binding proteins.**

(PDF)

**S4 Text. Testing the model with a synthetic dataset.**

(PDF)

**S5 Text. Mathematical proof about the robustness of model to the measured variables.**

(PDF)

## Author Contributions

**Conceptualization:** Yushan Qiu.

**Data curation:** Yushan Qiu.

**Formal analysis:** Liangjie Sun, Yushan Qiu, Pu Zhao, Quan Zou.

**Funding acquisition:** Yushan Qiu.

**Investigation:** Yushan Qiu.

**Methodology:** Liangjie Sun, Yushan Qiu, Quan Zou.

**Project administration:** Wai-Ki Ching.

**Software:** Liangjie Sun, Quan Zou.

**Supervision:** Yushan Qiu, Wai-Ki Ching, Quan Zou.

**Validation:** Liangjie Sun, Yushan Qiu, Pu Zhao.

**Writing – original draft:** Liangjie Sun.

**Writing – review & editing:** Liangjie Sun, Yushan Qiu, Wai-Ki Ching, Pu Zhao, Quan Zou.

## References

1. Yang J, Weinberg RA. Epithelial-mesenchymal transition: at the crossroads of development and tumor metastasis. *Developmental Cell*. 2008 Jun; 14(6):818–829. <https://doi.org/10.1016/j.devcel.2008.05.009> PMID: 18539112
2. Thiery JP, Acloque H, Huang RYJ, Nieto MA. Epithelial-mesenchymal transitions in development and disease. *Cell*. 2009 Nov; 139(5):871–890. <https://doi.org/10.1016/j.cell.2009.11.007> PMID: 19945376
3. Craene BD, Bex G. Regulatory networks defining EMT during cancer initiation and progression. *Nature Reviews Cancer*. 2013 Jan; 13(2):97–110. <https://doi.org/10.1038/nrc3447> PMID: 23344542
4. Mani SA, Yang J, Brooks M, Schwanger G, Zhou A, Miura N, et al. Mesenchyme Forkhead 1 (FOXC2) plays a key role in metastasis and is associated with aggressive basal-like breast cancers. *Proceedings of the National Academy of Sciences*. 2007 Jun; 104(24):10069–10074. <https://doi.org/10.1073/pnas.0703900104>
5. Bracken CP, Gregory PA, Kolesnikoff N, Bert AG, Wang J, Shannon MF, et al. A double-negative feedback loop between ZEB1-SIP1 and the microRNA-200 family regulates epithelial-mesenchymal transition. *Cancer Research*. 2008 Oct; 68(19):7846–7854. <https://doi.org/10.1158/0008-5472.CAN-08-1942> PMID: 18829540
6. Warzecha CC, Jiang P, Amirkian K, Dittmar KA, Lu H, Shen S, et al. An ESRP-regulated splicing programme is abrogated during the epithelial-mesenchymal transition. *The EMBO Journal*. 2010 Aug; 29(19):3286–3300. <https://doi.org/10.1038/emboj.2010.195> PMID: 20711167
7. Brown RL, Reinke LM, Damerow MS, Perez D, Chodosh LA, Yang J, et al. CD44 splice isoform switching in human and mouse epithelium is essential for epithelial-mesenchymal transition and breast cancer progression. *The Journal of Clinical Investigation*. 2011 Feb; 121(3):1064–1074. <https://doi.org/10.1172/JCI44540> PMID: 21393860
8. Shapiro IM, Cheng AW, Flytzanis NC, Balsamo M, Condeelis JS, Oktay MH, et al. An EMT-driven alternative splicing program occurs in human breast cancer and modulates cellular phenotype. *PLoS Genetics*. 2011 Aug; 7(8):e1002218. <https://doi.org/10.1371/journal.pgen.1002218> PMID: 21876675
9. Reinke LM, Xu Y, Cheng C. Snail represses the splicing regulator epithelial splicing regulatory protein 1 to promote epithelial-mesenchymal transition. *Journal of Biological Chemistry*. 2012 Oct; 287(43):36435–36442. <https://doi.org/10.1074/jbc.M112.397125> PMID: 22961986
10. Yang Y, Park JW, Bebee TW, Warzecha CC, Guo Y, Shang X, Xing Y, et al. Determination of a comprehensive alternative splicing regulatory network and combinatorial regulation by key factors during the epithelial-to-mesenchymal transition. *Molecular and Cellular Biology*. 2016 May; 36(11):1704–1719. <https://doi.org/10.1128/MCB.00019-16> PMID: 27044866
11. Wang ET, Sandberg R, Luo S, Khrebukova I, Zhang L, Mayr C, et al. Alternative isoform regulation in human tissue transcriptomes. *Nature*. 2008 Nov; 456(7221):470–476. <https://doi.org/10.1038/nature07509> PMID: 18978772
12. Nilsen TW, Graveley BR. Expansion of the eukaryotic proteome by alternative splicing. *Nature*. 2010 Jan; 463(7280):457–463. <https://doi.org/10.1038/nature08909> PMID: 20110989
13. Fu XD, Ares M. Context-dependent control of alternative splicing by RNA-binding proteins. *Nature Reviews Genetics*. 2014 Aug; 15(10):689–701. <https://doi.org/10.1038/nrg3778> PMID: 25112293
14. Damianov A, Ying Y, Lin CH, Lee JA, Tran D, Vashisht AA, et al. Rbfox proteins regulate splicing as part of a large multiprotein complex LASR. *Cell*. 2016 Apr; 165(3):606–619. <https://doi.org/10.1016/j.cell.2016.03.040> PMID: 27104978
15. Ying Y, Wang XJ, Vuong CK, Lin CH, Damianov A, Black DL. Splicing activation by Rbfox requires self-aggregation through its tyrosine-rich domain. *Cell*. 2017 Jul; 170(2):312–323. <https://doi.org/10.1016/j.cell.2017.06.022> PMID: 28708999

16. Barash Y, Calarco JA, Gao W, Pan Q, Wang X, Shai O, et al. Deciphering the splicing code. *Nature*. 2010 May; 465(7294):53–59. <https://doi.org/10.1038/nature09000> PMID: 20445623
17. Qiu Y, Ching WK, Zou Q. Prediction of rna-binding protein and alternative splicing event associations during epithelial-mesenchymal transition based on inductive matrix completion. *Briefings in Bioinformatics*. 2021 Sept; 22(5):bbaa440. <https://doi.org/10.1093/bib/bbaa440> PMID: 33517359
18. Qiu Y, Ching WK, Zou Q. Matrix factorization-based data fusion for the prediction of RNA-binding proteins and alternative splicing event associations during epithelial-mesenchymal transition. *Briefings in Bioinformatics*. 2021 Nov; 22(6):bbab332. <https://doi.org/10.1093/bib/bbab332> PMID: 34410342
19. Bendall SC, Davis KL, Amir ED, Tadmor MD, Simonds EF, Chen TJ, et al. Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. *Cell*. 2014 Apr; 157(3):714–725. <https://doi.org/10.1016/j.cell.2014.04.005> PMID: 24766814
20. Sun X, Zhang J, Nie Q. Inferring latent temporal progression and regulatory networks from cross-sectional transcriptomic data of cancer samples. *PLoS Computational Biology*. 2021 Mar; 17(3):e1008379. <https://doi.org/10.1371/journal.pcbi.1008379> PMID: 33667222
21. Qiu Y, Lyu J, Dunlap M, Harvey SE, Cheng C. A combinatorially regulated RNA splicing signature predicts breast cancer EMT states and patient survival. *RNA*. 2020 May; 26(9):1257–1267. <https://doi.org/10.1261/na.074187.119> PMID: 32467311
22. Rossell D, Attolini CSO, Kroiss M, Stöcker A. Quantifying alternative splicing from paired-end RNA-sequencing data. *The Annals of Applied Statistics*. 2014 Mar; 8(1):309. <https://doi.org/10.1214/13-aos687> PMID: 24795787
23. Katz Y, Wang ET, Airoidi EM, Burge CB. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nature methods*. 2010 Nov; 7(12):1009–1015. <https://doi.org/10.1038/nmeth.1528> PMID: 21057496
24. Van Nostrand EL, Freese P, Pratt GA, Wang X, Wei X, Xiao R, et al. A large-scale binding and functional map of human RNA-binding proteins. *Nature*. 2020 Jul; 583(7818):711–719. <https://doi.org/10.1038/s41586-020-2077-3> PMID: 32728246
25. Gerstberger S, Hafner M, Tuschl T. A census of human RNA-binding proteins. *Nature Reviews Genetics*. 2014 Nov; 15(12):829–845. <https://doi.org/10.1038/nrg3813> PMID: 25365966
26. Alves F, Dilão R. A simple framework to describe the regulation of gene expression in prokaryotes. *Comptes Rendus Biologies*. 2005 May; 328(5):429–444. <https://doi.org/10.1016/j.crv.2005.01.009> PMID: 15948632
27. Chan TE, Stumpf MPH, Babbie AC. Gene regulatory network inference from single-cell data using multivariate information measures. *Cell Systems*. 2017 Sept; 5(3):251–267. <https://doi.org/10.1016/j.cels.2017.08.014> PMID: 28957658
28. Seeger M, Steinke F, Tsuda K. Bayesian inference and optimal design in the sparse linear model. *Artificial Intelligence and Statistics*. 2007;444–451.
29. Hodson DJ, Janas ML, Galloway A, Bell SE, Andrews S, Li CM, et al. Deletion of the RNA-binding proteins ZFP36L1 and ZFP36L2 leads to perturbed thymic development and T lymphoblastic leukemia. *Nature Immunology*. 2010 Jul; 11(8):717–724. <https://doi.org/10.1038/ni.1901> PMID: 20622884
30. Cai Y, Ruan J, Yao X, Zhao L, Wang B. MicroRNA-187 modulates epithelial-mesenchymal transition by targeting PTRF in non-small cell lung cancer. *Oncology Reports*. 2017 Apr; 37(5):2787–2794. <https://doi.org/10.3892/or.2017.5548> PMID: 28393200
31. Georgiadis A, Tschernutter M, Bainbridge JWB, Balaggan KS, Mowat F, West EL, et al. The tight junction associated signalling proteins ZO-1 and ZONAB regulate retinal pigment epithelium homeostasis in mice. *PLoS One*. 2010 Sept; 5(12):e15730. <https://doi.org/10.1371/journal.pone.0015730> PMID: 21209887
32. Qiu Y, Jiang H, Ching WK, Ng MK. On predicting epithelial mesenchymal transition by integrating RNA-binding proteins and correlation data via  $L_{1/2}$ -regularization method. *Artificial Intelligence in Medicine*. 2019 Apr; 95:96–103. <https://doi.org/10.1016/j.artmed.2018.09.005> PMID: 30352711
33. Cieply B, Park JW, Nakauka-Ddamba A, Bebee TW, Guo Y, Shang X, et al. Multiphasic and dynamic changes in alternative splicing during induction of pluripotency are coordinated by numerous RNA-binding proteins. *Cell Reports*. 2016 Apr; 15(2):247–255. <https://doi.org/10.1016/j.celrep.2016.03.025> PMID: 27050523
34. Yan W, Deng Y, Zhang Y, Luo J, Lu D, Wan Q, et al. DZIP1 promotes proliferation, migration, and invasion of oral squamous carcinoma through the GLI1/3 pathway. *Translational Oncology*. 2019 Nov; 12(11):1504–1515. <https://doi.org/10.1016/j.tranon.2019.07.005> PMID: 31450126
35. Wang J, Huang Q, Liu ZP, Wang Y, Wu LY, Chen L, et al. NOA: a novel Network Ontology Analysis method. *Nucleic Acids Research*. 2011 Jul; 39(13):e87–e87. <https://doi.org/10.1093/nar/gkr251> PMID: 21543451

36. Pan QH, Fan YH, Wang YZ, Li DM, Hu CE, Li RX. Long noncoding RNA NNT-AS1 functions as an oncogene in breast cancer via repressing ZFP36 expression. *Journal of Biological Regulators and Homeostatic Agents*. 2020 May; 34(3):795–805. PMID: [32691576](#)
37. Seetharaman S, Flemyng E, Shen J, Conte MR, Ridley AJ. The RNA-binding protein LARP4 regulates cancer cell migration and invasion. *Cytoskeleton*. 2016 Sept; 73(11):680–690. <https://doi.org/10.1002/cm.21336> PMID: [27615744](#)
38. De Boeck M, Cui C, Mulder AA, Jost CR, Ikeno S, Ten Dijke P. Smad6 determines BMP-regulated invasive behaviour of breast cancer cells in a zebrafish xenograft model. *Scientific Reports*. 2016 Apr; 6(1):1–10. <https://doi.org/10.1038/srep24968>
39. <https://www.news.com.au/technology/science/human-body/genetic-research-gives-new-hope-for-sufferers-of-aggressive-breast-cancers/news-story/821fc417aae511dd018df61967d0ea37#:text=An>
40. Wang Z, Liu N, Shi S, Liu S, Lin H. The role of PIWIL4, an argonaute family protein, in breast cancer. *Journal of Biological Chemistry*. 2016 May; 291(20):10646–10658. <https://doi.org/10.1074/jbc.M116.723239> PMID: [26957540](#)
41. Heng ZSL, Lee JY, Subhramanyam CS, Wang C, Thanga LZ, Hu Q. The role of 17 $\beta$ -estradiol-induced upregulation of Piwi-like 4 in modulating gene expression and motility in breast cancer cells. *Oncology Reports*. 2018 Aug; 40(5):2525–2535. <https://doi.org/10.3892/or.2018.6676> PMID: [30226541](#)
42. Li AM, Ducker GS, Li Y, Seoane JA, Xiao Y, Melemenidis S, et al. Metabolic profiling reveals a dependency of human metastatic breast cancer on mitochondrial serine and one-carbon unit metabolism. *Molecular Cancer Research*. 2020 Apr; 18(4):599–611. <https://doi.org/10.1158/1541-7786.MCR-19-0606> PMID: [31941752](#)
43. Datta A, Deng S, Gopal V, Yap KCH, Halim CE, Lye ML, et al. Cytoskeletal dynamics in epithelial-mesenchymal transition: Insights into therapeutic targets for cancer metastasis. *Cancers*. 2021 Apr; 13(8):1882. <https://doi.org/10.3390/cancers13081882> PMID: [33919917](#)
44. Zhang H, Ma Y, Zhang S, Liu H, He H, Li N, et al. Involvement of Ras GTPase-activating protein SH3 domain-binding protein 1 in the epithelial-to-mesenchymal transition-induced metastasis of breast cancer cells via the Smad signaling pathway. *Oncotarget*. 2015 Jul; 6(19):17039. <https://doi.org/10.18632/oncotarget.3636> PMID: [25962958](#)
45. Zhao M, Kong L, Liu Y, Qu H. dbEMT: an epithelial-mesenchymal transition associated gene resource. *Scientific Reports*. 2015 Jun; 5(1):1–14. <https://doi.org/10.1038/srep11459> PMID: [26099468](#)
46. Bader GD, Hogue CWV. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics*. 2003 Jan; 4(1):1–27. <https://doi.org/10.1186/1471-2105-4-2> PMID: [12525261](#)
47. Zhitnyak IY, Rubtsova SN, Litovka NI, Gloushankova NA. Early events in actin cytoskeleton dynamics and E-cadherin-mediated cell-cell adhesion during epithelial-mesenchymal transition. *Cells*. 2020 Feb; 9(3):578. <https://doi.org/10.3390/cells9030578> PMID: [32121325](#)
48. Chen A, Beetham H, Black MA, Priya R, Telford BJ, Guest J, et al. E-cadherin loss alters cytoskeletal organization and adhesion in non-malignant breast cells but is insufficient to induce an epithelial-mesenchymal transition. *BMC Cancer*. 2014 Jul; 14(1):1–14. <https://doi.org/10.1186/1471-2407-14-552> PMID: [25079037](#)
49. Shankar J, Nabi IR. Actin cytoskeleton regulation of epithelial mesenchymal transition in metastatic cancer cells. *PLoS One*. 2015 Mar; 10(3):e0119954. <https://doi.org/10.1371/journal.pone.0119954> PMID: [25756282](#)
50. Feng XJ, Pan Q, Wang SM, Pan YC, Wang Q, Zhang HH, et al. MAP4K4 promotes epithelial-mesenchymal transition and metastasis in hepatocellular carcinoma. *Tumor Biology*. 2016 Mar; 37(8):11457–11467. <https://doi.org/10.1007/s13277-016-5022-1> PMID: [27010469](#)
51. Bambang I, Xu S, Zhou J, Salto-Tellez M, Sethi SK, Zhang D. Overexpression of endoplasmic reticulum protein 29 regulates mesenchymal-epithelial transition and suppresses xenograft tumor growth of invasive breast cancer cells. *Laboratory Investigation*. 2009 Sept; 89:1229–1242. <https://doi.org/10.1038/labinvest.2009.87> PMID: [19770839](#)
52. Zhang J, Shao X, Sun H, Liu K, Ding Z, Chen J, et al. NUMB negatively regulates the epithelial-mesenchymal transition of triple-negative breast cancer by antagonizing Notch signaling. *Oncotarget*. 2016 Spet; 7(38):61036. <https://doi.org/10.18632/oncotarget.11062> PMID: [27506933](#)
53. Huang CR, Lee CT, Chang KY, Chang WC, Liu YW, Lee JC, et al. Down-regulation of ARNT promotes cancer metastasis by activating the fibronectin/integrin  $\beta$ 1/FAK axis. *Oncotarget*. 2015 May; 6(13):11530–11546. <https://doi.org/10.18632/oncotarget.3448> PMID: [25839165](#)
54. Zhong C, Zuo Z, Ji Q, Feng D. P120ctn may participate in epithelial-mesenchymal transition in OSCC. *Indian Journal of Cancer*. 2016 Apr; 53(1):20–24. <https://doi.org/10.4103/0019-509X.180821> PMID: [27146732](#)

55. Yang J, Zhang L, Jiang Z, Ge C, Zhao F, Jiang J, et al. TCF12 promotes the tumorigenesis and metastasis of hepatocellular carcinoma via upregulation of CXCR4 expression. *Theranostics*. 2019 Aug; 9(20):5810–5827. <https://doi.org/10.7150/thno.34973> PMID: 31534521
56. Wang Y, Li N, Zheng Y, Wang A, Yu C, Song Z, et al. KIAA1217 Promotes Epithelial-Mesenchymal Transition and Hepatocellular Carcinoma Metastasis by Interacting with and Activating STAT3. *International Journal of Molecular Sciences*. 2021 Dec; 23(1):104. <https://doi.org/10.3390/ijms23010104> PMID: 35008530
57. Yuan J, Xing H, Li Y, Song Y, Zhang N, Xie M, et al. EPB41 suppresses the Wnt/ $\beta$ -catenin signaling in non-small cell lung cancer by sponging ALDOC. *Cancer Letters*. 2021 Feb; 499:255–264. <https://doi.org/10.1016/j.canlet.2020.11.024> PMID: 33242559
58. Hu X, Harvey SE, Zheng R, Lyu J, Grzeskowiak CL, Powell E, et al. The RNA-binding protein AKAP8 suppresses tumor metastasis by antagonizing EMT-associated alternative splicing. *Nature Communications*. 2020 Jan; 11(1):1–15. <https://doi.org/10.1038/s41467-020-14304-1> PMID: 31980632
59. Tao J, Sun D, Dong L, Zhu H, Hou H. Advancement in research and therapy of NF1 mutant malignant tumors. *Cancer Cell International*, 2020 Oct; 20(1):1–8. <https://doi.org/10.1186/s12935-020-01570-8> PMID: 33061844
60. Wang K, Ji W, Yu Y, Li Z, Niu X, Xia W, et al. FGFR1-ERK1/2-SOX2 axis promotes cell proliferation, epithelial-mesenchymal transition, and metastasis in FGFR1-amplified lung cancer. *Oncogene*. 2018 Jun; 37(39):5340–5354. <https://doi.org/10.1038/s41388-018-0311-3> PMID: 29858603
61. Liu J, Bang AG, Kintner C, Orth AP, Chanda SK, Ding S, et al. Identification of the Wnt signaling activator leucine-rich repeat in Flightless interaction protein 2 by a genome-wide functional analysis. *Proceedings of the National Academy of Sciences*. 2005 Jan; 102(6):1927–1932. <https://doi.org/10.1073/pnas.0409472102> PMID: 15677333
62. Zhang J, Tian XJ, Xing J. Signal transduction pathways of EMT induced by TGF- $\beta$ , SHH, and WNT and their crosstalks. *Journal of Clinical Medicine*. 2016 Mar; 5(4):41. <https://doi.org/10.3390/jcm5040041> PMID: 27043642
63. Ackermann A, Brieger A. The Role of Nonerythroid Spectrin *all* in Cancer. *Journal of Oncology*. 2019 May; 2019:7079604. <https://doi.org/10.1155/2019/7079604> PMID: 31186638
64. Song Y, Zheng S, Wang J, Long H, Fang L, Wang G, et al. Hypoxia-induced PLOD2 promotes proliferation, migration and invasion via PI3K/Akt signaling in glioma. *Oncotarget*. 2017 Jun; 8(26):41947–41962. <https://doi.org/10.18632/oncotarget.16710> PMID: 28410212
65. Strippoli R, Benedicto I, Perez Lozano ML, Pellinen T, Sandoval P, Lopez-Cabrera M, et al. Inhibition of transforming growth factor-activated kinase 1 (TAK1) blocks and reverses epithelial to mesenchymal transition of mesothelial cells. *PloS One*. 2012 Feb; 7(2):e31492. <https://doi.org/10.1371/journal.pone.0031492> PMID: 22384029
66. Tang L, Zhao P, Kong D. Muscblind-like 1 destabilizes Snail mRNA and suppresses the metastasis of colorectal cancer cells via the Snail/E-cadherin axis. *International Journal of Oncology*. 2019 Jan; 54(3):955–965. <https://doi.org/10.3892/ijo.2019.4691> PMID: 30664186
67. Li X, Jiang F, Wang X, Gu X. SPAG9 regulates HEF1 expression and drives EMT in bladder transitional cell carcinoma via rac1 signaling pathway. *American Journal of Cancer Research*. 2018 Dec; 8(12):2467–2480. PMID: 30662804
68. Cusseddu R, Robert A, Côté JF. Strength through unity: The power of the mega-scaffold macf1. *Frontiers in Cell and Developmental Biology*. 2021 Mar; 9:641727. <https://doi.org/10.3389/fcell.2021.641727> PMID: 33816492
69. Lu H, Liu J, Liu S, Zeng J, Ding D, Carstens RP, et al. Exo70 isoform switching upon epithelial-mesenchymal transition mediates cancer cell invasion. *Developmental Cell*. 2013 Dec; 27(5):560–573. <https://doi.org/10.1016/j.devcel.2013.10.020> PMID: 24331928
70. Risolino M, Mandia N, Iavarone F, Dardaei L, Longobardi E, Fernandez S, et al. Transcription factor PREP1 induces EMT and metastasis by controlling the TGF- $\beta$ -SMAD3 pathway in non-small cell lung adenocarcinoma. *Proceedings of the National Academy of Sciences*. 2014 Aug; 111(36):E3775–E3784. <https://doi.org/10.1073/pnas.1407074111> PMID: 25157139
71. Volakis LI, Li R, Ackerman WE IV, Mihai C, Bechel M, Summerfield TL, et al. Loss of myoferlin redirects breast cancer cell motility towards collective migration. *PloS One*. 2014 Feb; 9(2):e86110. <https://doi.org/10.1371/journal.pone.0086110> PMID: 24586247
72. Li R, Ackerman WE IV, Mihai C, Volakis LI, Ghadiali S, Kniss DA. Myoferlin depletion in breast cancer cells promotes mesenchymal to epithelial shape change and stalls invasion. *PloS One*. 2012 Jun; 7(6):e39766. <https://doi.org/10.1371/journal.pone.0039766> PMID: 22761893
73. Zhou W, Thiery JP. Loss of Git2 induces epithelial-mesenchymal transition by miR146a-Cnot6L-controlled expression of Zeb1. *Journal of Cell Science*. 2013 Jun; 126(12):2740–2746. PMID: 23591815

74. Timmerman LA, Grego-Bessa J, Raya A, Bertrán E, Pérez-Pomares JM, Díez J, et al. Notch promotes epithelial-mesenchymal transition during cardiac development and oncogenic transformation. *Genes & Development*. 2004; 18(1):99–115. <https://doi.org/10.1101/gad.276304> PMID: 14701881
75. Xia P, Gu R, Zhang W, Sun YF. lncRNA CEBPA-AS1 overexpression inhibits proliferation and migration and stimulates apoptosis of OS cells via notch signaling. *Molecular Therapy-Nucleic Acids*. 2020 Mar; 19:1470–1481. <https://doi.org/10.1016/j.omtn.2019.10.017> PMID: 32160715
76. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. CytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Systems Biology*. 2014 Dec; 8(4):1–7. <https://doi.org/10.1186/1752-0509-8-S4-S11> PMID: 25521941
77. Xu Y, Gao XD, Lee JH, Huang H, Tan H, Ahn J, et al. Cell type-restricted activity of hnRNPM promotes breast cancer metastasis via regulating alternative splicing. *Genes & Development*. 2014 May; 28(11):1191–1203. <https://doi.org/10.1101/gad.241968.114> PMID: 24840202
78. Zhao P, Xu Y, Wei Y, Qiu Q, Chew TL, Kang Y, et al. The CD44s splice isoform is a central mediator for invadopodia activity. *Journal of Cell Science*. 2016 Apr; 129(7):1355–1365. <https://doi.org/10.1242/jcs.171959> PMID: 26869223
79. Warzecha CC, Shen S, Xing Y, Carstens RP. The epithelial splicing factors ESRP1 and ESRP2 positively and negatively regulate diverse types of alternative splicing events. *RNA Biology*. 2009 Nov; 6(5):546–562. <https://doi.org/10.4161/rna.6.5.9606> PMID: 19829082
80. Ishii H, Saitoh M, Sakamoto K, Kondo T, Katoh R, Tanaka S, et al. Epithelial splicing regulatory proteins 1 (ESRP1) and 2 (ESRP2) suppress cancer cell motility via different mechanisms. *Journal of Biological Chemistry*. 2014 Oct; 289(40):27386–27399. <https://doi.org/10.1074/jbc.M114.589432> PMID: 25143390