

# Causal Inference Network of Genes Related with Bone Metastasis of Breast Cancer and Osteoblasts Using Causal Bayesian Networks

Sung Bae Park<sup>1</sup>, Chun Kee Chung<sup>2,\*</sup>, Efrain Gonzalez<sup>3</sup>, Changwon Yoo<sup>3,\*</sup>

<sup>1</sup>Department of Neurosurgery, Seoul National University Boramae Medical Center, Seoul;

<sup>2</sup>Department of Neurosurgery, Seoul National University Hospital, Seoul National University College of Medicine, Clinical Research Institute, Seoul, Korea

<sup>3</sup>Department of Biostatistics, Robert Stempel College of Public Health and Social Work, Florida International University, Miami, FL, USA

## Corresponding author

Changwon Yoo

Department of Biostatistics, Robert Stempel  
College of Public Health and Social Work,  
Florida International University, 11200 SW 8th  
Street AHC5, Miami, FL 33199, USA  
Tel: +1-305-348-4906  
Fax: +1-305-348-4901  
E-mail: cyoo@fiu.edu

Chun Kee Chung

Department of Neurosurgery, Seoul National  
University Hospital, 101 Daehak-ro, Jongno-gu,  
Seoul 03080, Korea  
Tel: +82-2-2072-2352  
Fax: +82-2-744-8459  
E-mail: chungc@snu.ac.kr

Received: October 7, 2018

Revised: October 29, 2018

Accepted: November 2, 2018

No potential conflict of interest relevant to this article was reported.

\*Changwon Yoo and Chun Kee Chung contributed equally to this work and should be considered co-corresponding authors.

This study was supported by a grant from SNUH research fund (03-2015-0180).

Copyright © 2018 The Korean Society for Bone and Mineral Research

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Background:** The causal networks among genes that are commonly expressed in osteoblasts and during bone metastasis (BM) of breast cancer (BC) are not well understood. Here, we developed a machine learning method to obtain a plausible causal network of genes that are commonly expressed during BM and in osteoblasts in BC. **Methods:** We selected BC genes that are commonly expressed during BM and in osteoblasts from the Gene Expression Omnibus database. Bayesian Network Inference with Java Objects (Banjo) was used to obtain the Bayesian network. Genes registered as BC related genes were included as candidate genes in the implementation of Banjo. Next, we obtained the Bayesian structure and assessed the prediction rate for BM, conditional independence among nodes, and causality among nodes. Furthermore, we reported the maximum relative risks (RRs) of combined gene expression of the genes in the model. **Results:** We mechanistically identified 33 significantly related and plausibly involved genes in the development of BC BM. Further model evaluations showed that 16 genes were enough for a model to be statistically significant in terms of maximum likelihood of the causal Bayesian networks (CBNs) and for correct prediction of BM of BC. Maximum RRs of combined gene expression patterns showed that the expression levels of *UBIAD1*, *HEBP1*, *BTNL8*, *TSP0*, *PSAT1*, and *ZFP36L2* significantly affected development of BM from BC. **Conclusions:** The CBN structure can be used as a reasonable inference network for accurately predicting BM in BC.

**Key Words:** Bayes theorem · Breast neoplasms · Neoplasm metastasis · Osteoblasts

## INTRODUCTION

Bone is a frequent site for primary tumor cells to spread and more than 600,000 people every year suffer from bone metastasis (BM) in the US.[1] Breast, prostate and lung cancers commonly lead to BM and more than 70% of patients with breast cancer (BC) and BM have a high rate of morbidity and mortality.[2-4] Patients with BM may suffer from a variety of skeletal-related complications (SRCs) such as bone pain, pathologic fractures, spinal cord compression, and difficulty to walk.[5] Additionally, once a patient with BM experiences a first SRC, the likelihood of experiencing a second SRC is greatly increased.[5] Studies regarding normal physiology

of bone metabolism, interactions between bone and advanced cancer, and genetics related with BM have been performed in the laboratory.[6-8] In addition, surgeries, radiation therapy and chemical therapy using target agents are performed in the clinical field.[9,10] However, despite the multidisciplinary approach, a large proportion of BM remains incurable and treatment strategies for BM usually have a palliative effect.[6] The decrease in effectiveness is due to fact that the pathophysiology of BM from BC is a multistep and complex process that consists of escapes of tumor cell from primary site, a crosstalk between disseminated BC cells and bone-derived molecules, leading to bone and reconstitution of secondary tumors at the bone.[11,12] Although novel therapies which target the pathways involved in BM are emerging, research which focuses on identifying key upstream regulators related with the molecular signaling pathway of BM remains clinically relevant to the prevention of BM.[11] Using statistical machine learning methods can help us find key upstream regulators from a causal network model inferred from clinical, genomic and environmental data related with BM. In turn, this can improve clinically preventive practices and reduce the adverse effect related with target therapy of BM.

Application of machine learning methods have been widely used to get the statistical relationship and causal networks from large and complicated health data.[13,14] As one of machine learning methods, causal Bayesian networks (CBN) have been used to learn causal network inferred from known collected genomic data.[15,16] It has been shown that the gene expression patterns could be influenced by multistep complex processes from primary site, through dissemination, into metastasis.[7,17] Therefore, the relationships of 1 or 2 genes or molecules cannot reflect how combination of many genes' expressions changes according to surrounding environment. We need to seek for the relationships from interactions from network of genes and further search for plausible causal interactions. Although studies have presented the gene changes according to BM of BC (BMBC), none of the studies developed research to build an integrated causal network consisting of gene signatures associated to BMBC.[17] The CBN analysis of dataset of the expressed genes in BMBC can provide insight into the causal relationship and upstream regulators governing BMBC. There are 3 specific niches in homing of tumor cells into bone: the endosteal niche, the haemato-

poietic stem cell niche and the vascular niche.[18] The endosteal niche related with osteoblast has a key role in the homing and adhesion of circulating tumor cells into bone and the clinically relevant and important network related with BM seems to be in the endosteal niche.[18,19] Therefore, we conducted a CBN analysis of microarray data from the Gene Expression Omnibus (GEO) to get a causal gene expression network consisting of genes related with BMBC and osteoblast for identifying the upstream regulators of BMBC.

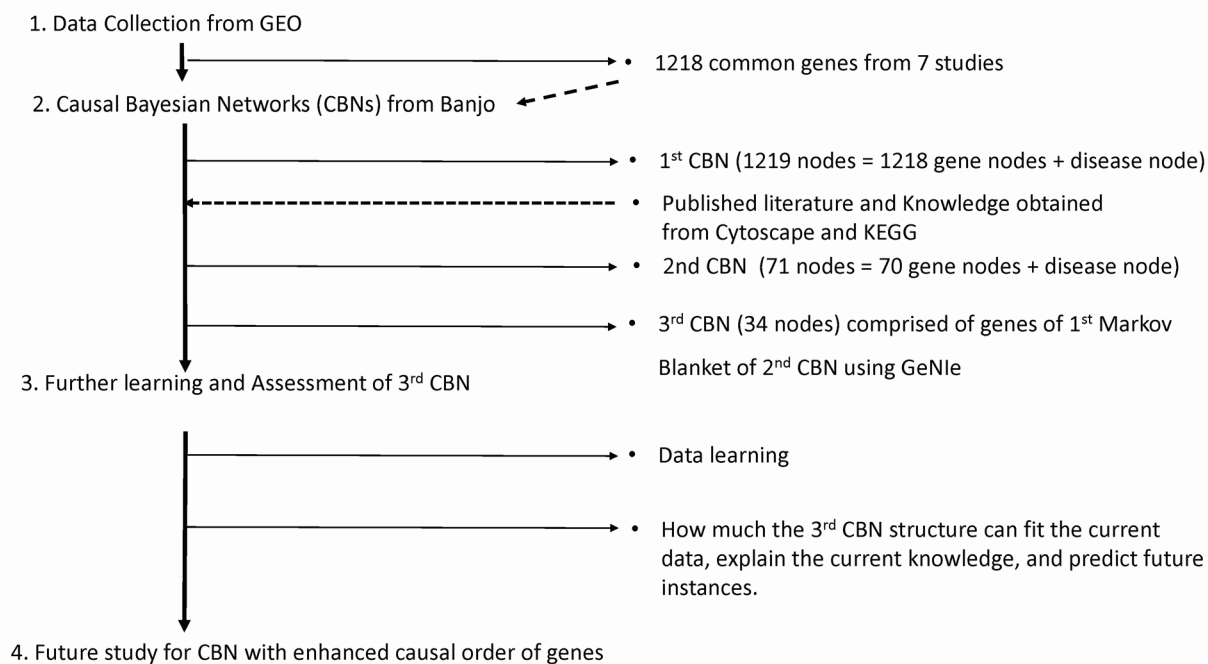
## METHODS

After collecting gene expression data from GEO following inclusion criteria, we cleaned, normalized, and discretized the gene expression levels. Then, we selected candidate genes from the list of genes that were common to the studies involving BMBC and osteoblast. In the next step, we learned CBN structure by combining the gene expression data with published literature and knowledge. We further learned CBN parameters (conditional probabilities) and evaluated the final model to see how well it fits the data (maximum likelihood and conditional independencies), compared it with current knowledge (published literature articles) and how well it predicts future instances (receiver operating characteristic [ROC] curve and relative risk [RR]) (Fig. 1).

We further describe each step that we summarized above in more detail:

### 1. Collecting data from GEO

We have retrieved microarray datasets from the GEO database of the National Center for Biotechnology Information (NCBI) of National Institute of Health (<https://www.ncbi.nlm.nih.gov/geo/>, GEO, NCBI accessed in November 2017).[20] We have used the following inclusion criteria: (1) datasets with the GEO series (GSE) of BMBC, BC without metastasis and osteoblast; (2) studies should be measuring gene expression of human (*homo sapiens*); (3) studies tissue extracted from metastatic bone and the cancerous region of breast and experimental human cell lines or normal bone tissue for osteoblast; and (4) BC without metastasis was diagnosed as breast ductal carcinoma without metastasis when the tissues were extracted by biopsy or surgery. As a result, there were seven studies that were iden-



**Fig. 1.** Outline of the study. GEO, Gene Expression Omnibus; KEGG, Kyoto Encyclopedia of Genes and Genomes; CBN, causal Bayesian network.

tified for this study (ten GSEs and 48 GEO sample) (Table 1).

## 2. Data mining and selection of candidate genes

In this section, we describe how we prepared the datasets that were collected from GEO and in the next section we describe how we learn causal network of genes commonly expressed in BMBC and osteoblasts using CBN.

### 1) Data cleaning, averaging, normalization and discretization (CANDi)

Because gene symbols in the GSEs (set of samples) are written by gene Identifier (ID) not gene name, we changed gene IDs in GSE into corresponding gene symbol in matched GPL (technology platform of microarray analysis used in study) (Cleaning). Then, if single gene had several raw expression values of gene, we transformed the raw values of each gene into a mean value of the raw values (Averaging). For meaningful comparison of gene expression level, the relative data from individual studies should be transformed into normalized data. We normalized the averaging data into z-scores and performed the normalization task on study-by-study basis (Normalization).[21] Bayesian Network Inference with Java Objects (Banjo) is one of computational modeling tools based on data-driven method and Banjo utilize Bayesian network frameworks to result in directed

inference network.[2,22] Because Banjo implement Bayesian Dirichlet equivalence scoring metric of data consisting of discrete variables, we transformed the normalized data to discrete values (Discretization). For each gene, all expressions z score values (z) were discretized as less than -1 ( $z < -1$ ), between -1 and 1 ( $-1 \leq z \leq 1$ ), and over 1 ( $z > 1$ ) were discretized as low, no change, and high, respectively. We sequentially performed the same CANDi process per each study.

### 2) Selection of candidate genes

We selected common genes there were all presented in GSE of osteoblast, normal bone tissue and BMBC. Among the common genes, if the genes simultaneously expressed in BC without metastasis such as genes expressed in BC in-situ, we excluded the genes in the study. As a result, ten GSEs (5 for osteoblast and 5 for BMBC) from seven studies and 1,218 genes were selected (Table 1). After collection of all data in ten GSEs, we prepared a dataset (denote as  $D^\circ$ ) that consists of variables that represent expression levels (low, no change, and high) of 1,218 genes and a variable called *group* that represents whether a subject has *osteoblast* or *BM*. In the final dataset  $D^\circ$ , there were 48 subjects (13 GSMS with *osteoblast* and 35 GSMS with *BMv*) (Table 1).

**Table 1.** Information of enrolled Gene Expression Omnibus Series experiments and Gene Expression Omnibus Simple Omnibus Format in Text format Sample files

References	Study title	GSE number	GSM number	GSM source	GSM gender and age	Gene Information related with bone metastasis
1	Endo-Munoz et al.[49]	19276	481115 481116 481117 481118 481119	Normal human bone	Female and 49 years Female and 75 years Male and 71 years Male and 56 years Male and 49 years	• Non available
2	Rajski et al.[40]	29033 29034 29035	719377 719381 719372 719376 719384 719387	Normal human osteo- blast	Non available	• An <b>IL-6</b> gene expression pattern induced by heterotypic interaction of breast cancer cells with osteoblasts <i>in vitro</i> is associated with a higher rate of bone metastasis <i>in vivo</i> .
3	Sadikovic et al.[50]	12885	288005 288006	Normal human osteo- blast	Healthy male donor Healthy male donor	• Non available
4	Zhang et al.[51] Xu et al.[52]	14017  14018	352100 352103 352105 352109 352117 352119 352123 352124 352126 352131 352144 352149 352151 352154 352155 352159 352163 352167	Metastatic bone speci- men of breast cancer	Human, other informa- tion is non available  Human, other informa- tion is non available	• A strong association between late-onset bone metastasis and c-Src activity (c-Src protein is encoded by <b>SRC gene</b> ) • c-Src supports cancer cell survival in the bone marrow microenvironment by facilitating CXCL12-CXCR4-AKT signaling and by conferring resistance to TRAIL (CXCL12, CXCR4, AKT and TRAIL are encoded by <b>CXC12, CXCR4, AKT1 and TRAIL genes</b> , respectively). • 14-3-3 $\zeta$ stabilizes Gli2 in breast cancer cells, and Gli2 partners with Smads to activate PTHrP and promote TGF- $\beta$ -induced bone metastasis (14-3-3 $\zeta$ , Gli2, PTHrP and TGF- $\beta$ are encoded by <b>YWHAZ, GLI2, SMAD, PTHLH and Tgfbr2 genes</b> , respectively).
5	Mourskaia et al.[53]	39494	970199 970200 970201 970203	Metastatic bone speci- men of breast cancer	Human, other informa- tion is non available	• ABCC5, an ATP-dependent transporter, was found to be overexpressed in breast cancer osseous metastases relative to primary breast tumors. <b>ABCC5 gene</b> expression in breast cancer cells is important for efficient osteoclast-mediated bone resorption. (ABCC5 gene is related with osteoclast not osteoblast).

(continued to the next page)

Table 1. Continued

References	Study title	GSE number	GSM number	GSM source	GSM gender and age	Gene Information related with bone metastasis
6 Kimbung et al.[54]	Expression data from fine-needle aspiration biopsies of breast cancer metastases from different anatomical sites	46141	1124888 1124890 1124904 1124927 1124953	Metastatic bone specimens of breast cancer	Human, other information is non available	Gene Information related with bone metastasis • CLDN2 was significantly up-regulated ( $P < 0.001$ ) in liver metastases compared to other metastatic sites.
7 Not published	Identification of genes associated with bone metastatic breast cancer cells	51232	1240799 1240800 1240801 1240803 1240804 1240805 1240806	Bone metastasis derivatives	Bone metastatic specimens induced by injection of MCF7	• MAF is a mediator of breast cancer bone metastasis (MAF gene encodes MAF mediator)

GSE, Gene Expression Omnibus Series experiments; GSM, Gene Expression Omnibus Simple Omnibus Format in Text format Sample file.

### 3. Learning CBN structure

The structure of CBN is directed acyclic graph (DAG) consisting of nodes and intervening arrows.[13] The arrows in DAG are interpreted in terms of probabilistic conditional independence between nodes.[23] First degree Markov Blanket (MB) of a variable  $X$  in CBN (denote as  $MB[X]$ ) is defined as set of variables that is the direct causes (parents) of  $X$  and direct effects (children) of  $X$  and direct causes (parents) of direct effects (children) of  $X$  (of course,  $X$  itself is excluded from  $MB[X]$ ). Second degree MB of  $X$ , third degree MB of  $X$ , etc., can be defined as  $MB(MB[X])$ ,  $MB(MB(MB[X]))$ , etc., respectively.

An example CBN structure is shown in Figure 2. In that structure, for example, the expression of NFKB2 influences the likelihood of the presence of BM which influences on expressions of MAP2K2 and TSO. In parallel, conditional independence between nodes shows as the probability about expression of MAP2K2 or TSO is not influenced by NFKB2 given information of BM. Also in Figure 2, first degree MB of MAP2K2,  $MB(MAP2K2)$ , is {BM, TSO} and second degree MB of MAP2K2,  $MB(MB[MAP2K2])$ , is {NFKB2, BM, TSO}.

We used Banjo,[24] a Bayesian network learning tool, to learn 2 CBNs, i.e., a CBN with all 1,218 candidate genes from

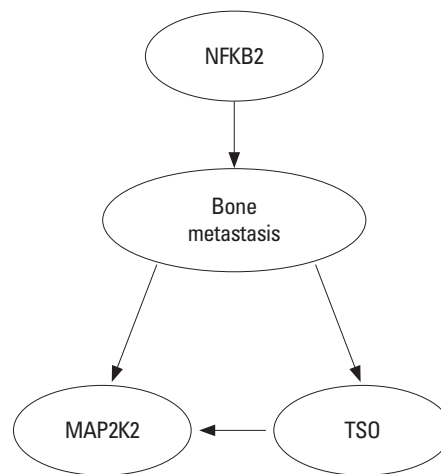


Fig. 2. An example of causal Bayesian networks structure. The expression of NFKB2 influences the likelihood of the presence of bone metastasis which influences on expressions of MAP2K2 and TSO. In parallel, conditional independence between nodes shows as the probability about expression of MAP2K2 or TSO is not influenced by NFKB2 given information of bone metastasis. The first degree Markov Blanket (MB) of MAP2K2,  $MB(MAP2K2)$ , is {Bone metastasis, TSO} and second degree MB of MAP2K2,  $MB(MB[MAP2K2])$ , is {NFKB2, Bone metastasis, TSO}.



all the studies and a CBN with 70 BC-Relevant Genes (BCRGs).

### 1) CBN with candidate genes

To learn the CBN with all 1,218 candidate genes from all the studies, we ran 18 independent runs of Banjo (3 independent 1 hr, 3 hr, 6 hr, 12 hr, 24 hr, and 36 hr runs; total of 246 hr runs) with the dataset  $D^\circ$  and by limiting maximum parents to be 5. Among 18 best log likelihood structures that were reported by each independent Banjo run, we chose the network with the highest log likelihood (denoted as  $S^\circ$ ; note that  $S^\circ$  includes 1,219-1,218 variables, each representing a gene expression and a variable named *group*, representing whether a subject has *osteoblast* or *BM*).

### 2) CBN with BCRGs

CluePedia, one of Cytoscape plugins, have been known as a provider for pathways, processes or disease related with gene, miRNA and protein in conjunction with ClueGo. [25] Using ClueGo and CluePedia Cytoscape plugin, we found 13 BCRGs (that represents genes having more close association to disease and acceptable prior knowledge) registered in kyoto encyclopedia of genes and genomes (KEGG). [26] To further analyze possible interactions of 13 BCRGs and 57 genes from second degree MB of *group* node in  $S^\circ$ , we again ran Banjo with a new dataset (denote as  $D^\star$ ) that combines the expressions of these 70 genes (13 BCRGs and 57 genes from second degree MB). The condition of setting files for the second Banjo analysis was same as previous condition except that we used  $D^\star$  with data of 71 variables extracted from data of 1,219 variables in  $D^\circ$  (48 subjects). Among 18 best log likelihood structures that were reported by each independent Banjo run, we chose the network with the highest log likelihood (denoted as  $S^\star$ ; note that  $S^\star$  includes 71-70 variables, each representing a gene expression and a variable named *group*, representing whether a subject has *osteoblast* or *BM*).

## 4. Learning causal Bayesian parameters

In the network with the highest log likelihood in the second Banjo analysis ( $S^\star$ ), we represent first degree MB of the *group* variable using Bayesian network graphical network interface (GeNIe; version 2.2.1; BayesFusion, Pittsburgh, PA, USA) to learn parameters (probabilities). Learning the parameters of the Bayesian structure was implemented using the data set with the dataset that was extracted from

the original dataset (48 observations). In addition, we investigated the relationships among nodes and the influence of status of BMBC on expressions of other genes in the Bayesian structure which comprised of 34 nodes within MB (group) of  $S^\star$  (denoted as  $S^*$ ; note that  $S^*$  includes 34–33 variables, each representing a gene expression and a variable named *group*, representing whether a subject has *osteoblast* or *BM*).

## 5. Validations

We further validated CBN structure  $S^*$  by how well  $S^*$  fits the data (maximum likelihood and conditional independencies), compared  $S^*$  with current knowledge (published literature articles) and how well  $S^*$  predicts future instances (ROC and RR). We searched for orders of variables with best probabilistic score using Markov chain over Monte Carlo (MCMC) simulation and compared the result with the order of nodes in Bayesian structure (call this order analysis). [27] The variables we used were direct cause (parent) and direct effect (children) genes and disease node that showed the strongest influence through conditional independency test (we refer these variables as *Validation Variables* and denote as  $V^*$ ). [28] We performed order analysis with 3 independent 1 hour runs and we compare the best order with  $S^*$  and current knowledge. Also we investigated degree of conditional independencies among variables in CBN structure  $S^*$  and check how well conditional independency relationships (d-separation and d-connectivity [29]) among variables. We also evaluated CBN structure  $S^*$  using leave one out cross test and the area under ROC (AUROC) using GeNIe. We obtained the prediction rate of BMBC of CBN structure  $S^*$  with the final dataset that includes the same number of variables in  $S^*$ , i.e., 34-33 variables, each representing a gene expression and a variable named *group*, representing whether a subject has *osteoblast* or *BM*, and 48 subjects. We calculate how much we can expect BM with the final dataset of BM if we have the information in *Validation Variables*  $V^*$ . We investigated the information of genes within in CBN structure  $S^*$  whether the structure reflected the current knowledge and previous studies using Cytoscape 3.6 [26] and searching articles in PubMed in NCBI ([www.ncbi.nlm.nih.gov/pubmed](http://www.ncbi.nlm.nih.gov/pubmed)).

In final process of validation, we investigated all joint distributions of genes in *Validation Variables*  $V^*$  given BMBC and also how all possible combination of gene expression

patterns predict BMBC and evaluated the minimum combination with maximum RR of combined gene expression of the genes in the final model and significant genes. We have used 12 genes in *Validation Variables*  $V^*$  and used CBN  $S^*$  and calculated all 16,777,215 ( $= \sum_{i=1}^{12} \binom{12}{i} 3^i = (3+1)^{12} - 1 = 4^{12} - 1$ ) different gene configurations  $g$  with the collected dataset and using SMILE library (<https://www.bayesfusion.com/smile-engine>) and C++ program and calculate the following:

$$P(D|V=g)$$

where  $D$  represents a subject has BMBC and  $V$  is any subset of  $V^*$ . Among the gene configurations  $g$  that predicts BMBC with high or low probability (i.e.,  $P(D|V=g) > 0.99999$  or  $P(D|V=g) < 1.0 \times 10^{-6}$ ).

To find the minimum set of combination of gene expression patterns that give us a maximum RR, we have calculated the following:

$$R = \underset{V}{\operatorname{argmax}} \frac{P(D|V = v)}{P(D|V = v')}$$

where,  $V$  represents any subset of  $V^*$ ,  $R$  represents a set of the minimum number of genes that maximizes the RR

term,  $v = \underset{g}{\operatorname{argmax}} P(D|V = g)$  and  $v' = \underset{g}{\operatorname{argmin}} P(D|V = g)$ . Note that  $v$  and  $v'$  represents 2 different gene expression patterns among the genes in  $R$  that maximizes and minimizes  $P(D|V)$  respectively. We report the top 4 RR that we calculated from the dataset.

## RESULTS

We report the 3 CBN structures, i.e., a CBN with candidate genes ( $S^\circ$ ), a CBN with BCRGs ( $S^\star$ ), and a CBN as the first degree MB of *group* variable in  $S^\star$  ( $S^*$ ) described in the method section. We also report results of their validations.

### 1. CBN with common genes of all studies

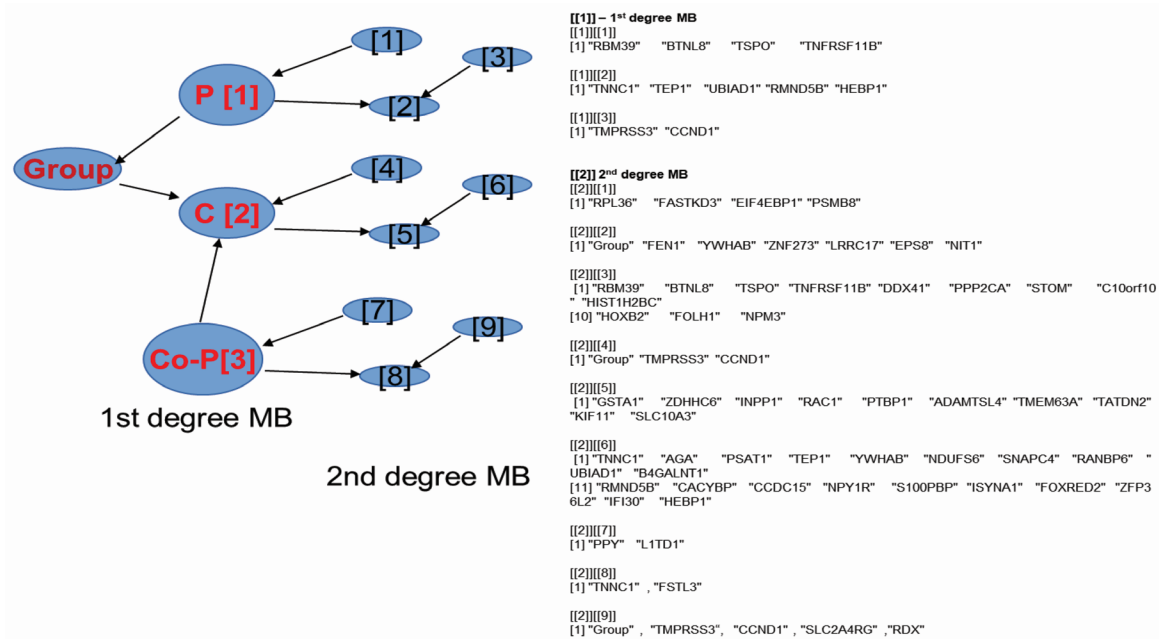
Among the 18 CBNs, the CBN with the best log likelihood score reported by Banjo (i.e.,  $\log P(D^\circ | S^\circ) = -49,091.645$  where  $S^\circ$  a CBN with 1,219 variables and  $D^\circ$  is the dataset with the same number of variables with 48 subjects) was significantly better fitting the data than the second best CBN (i.e.,  $\frac{P(D^\circ | S^{\circ'})}{P(D^\circ | S^\circ) + P(D^\circ | S^{\circ'})} > 99.999\%$  where  $S^{\circ'}$  is the second best CBN).  $S^\circ$  includes 1,218 gene expression and a variable called



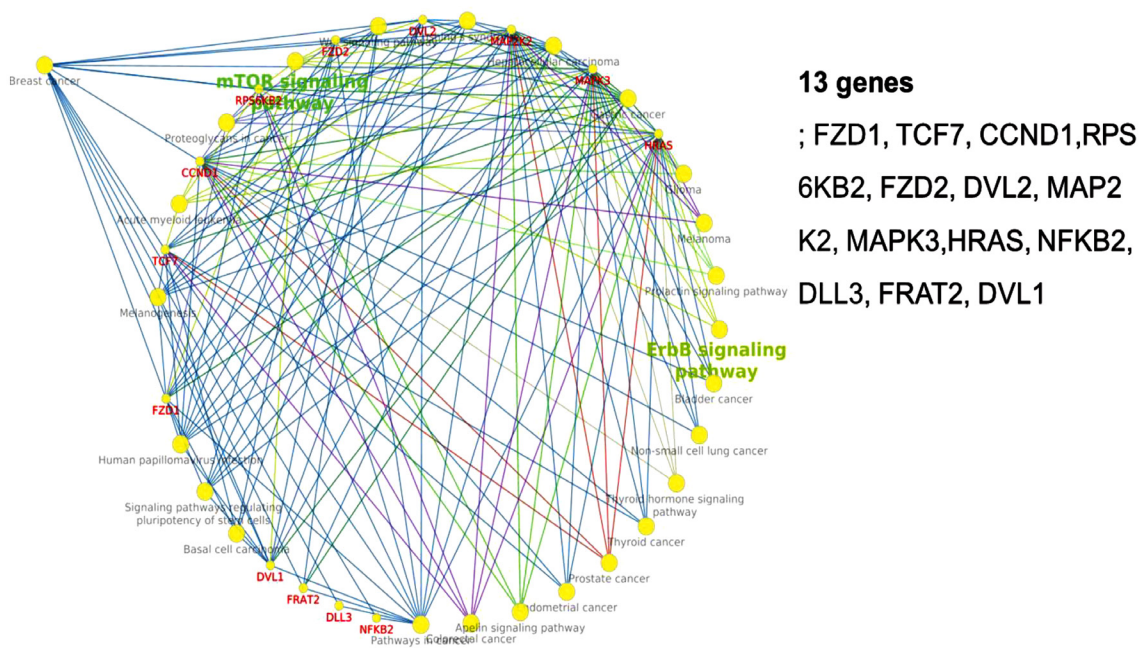
**Fig. 3.** Causal Bayesian network structure. Left figure shows the 1st causal Bayesian network structure with 1,219 node and complicated connection between them. Right figure show red colored group node and around connected genes. Group node represented osteoblast or bone metastasis of breast cancer.

group (representing whether a subject has *osteoblast* or *BM*) (Fig. 3). Among 1,219 variables, we selected 59 genes within second degree MB of *group* (Fig. 4). In addition, CluePedia, one of Cytoscape plugins, identified 13 BCRGs that

had common pathways – mammalian target of rapamycin signaling pathway, ErbB signaling pathway, signaling pathways regulating pluripotency of stem cell – in prostate cancers, thyroid cancer and non-small cell lung cancer (Fig. 5).



**Fig. 4.** Schematic description and genes in 2nd Markov blanket. This figure show the schematic description of relationship among group node, parent (P) node, child (C) node and co-parent (Co-P) node (left-side figure) and gene names (right-side figure) within 2nd degree Markov blanket (MB).



**Fig. 5.** Circular layout with breast cancer related genes, pathways, and diseases. Cytoscape with ClueGo and CluePedia showed us the circular structures with 13 breast cancer relevant genes, pathways, and other diseases as well as connections among them.



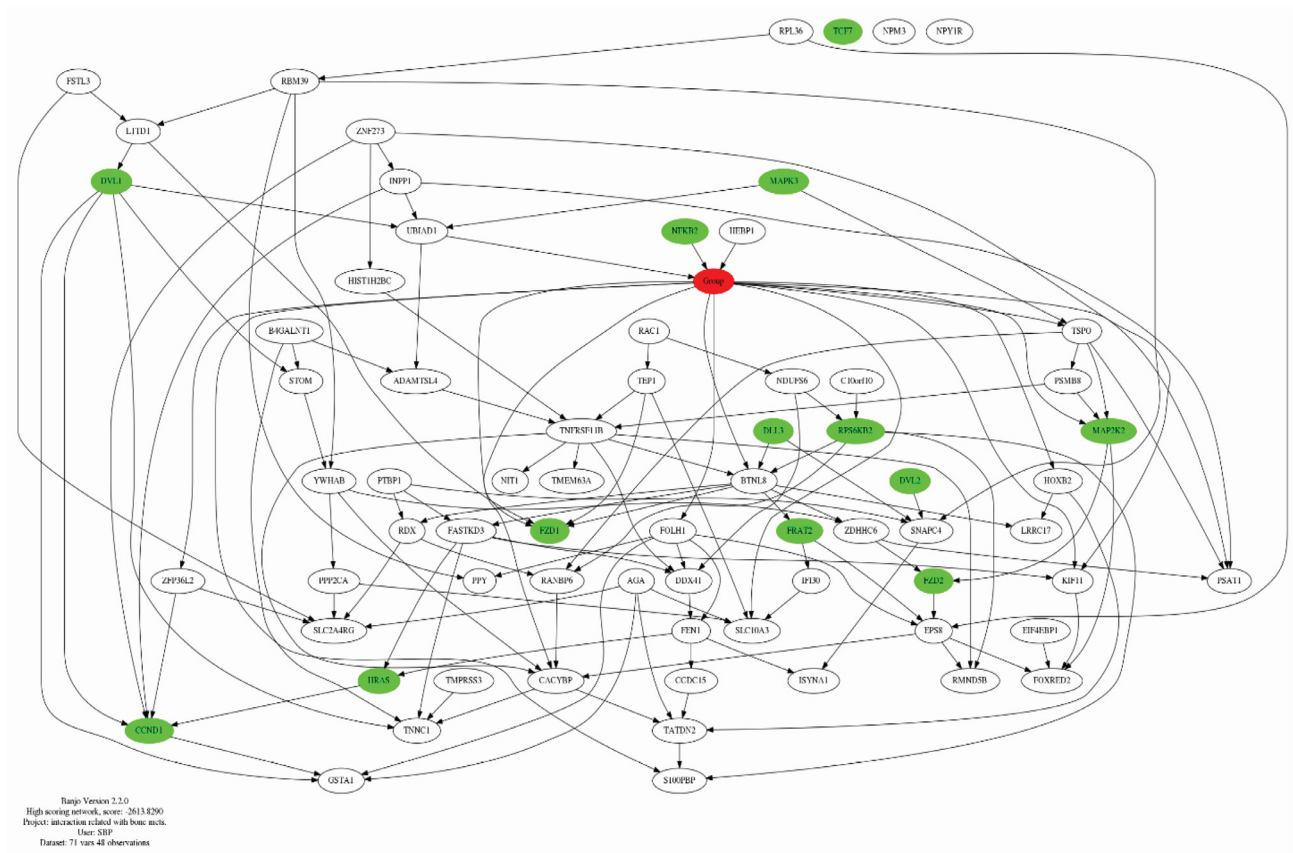
### 2. CBN with BCRGs

Among the 18 CBNs, the CBN with the best log likelihood score reported by Banjo (i.e.,  $\log P(D^* | S^*) = -2,613.829$  where  $S^*$  a CBN with 71 variables and  $D^*$  is the dataset with the same number of variables with 48 subjects) was significantly better fitting the data than the second best CBN (i.e.,  $\frac{P(D^* | S^*)}{P(D^* | S^*) + P(D^* | S')}$  > 99.999% where  $S'$  is the second best CBN) (Fig. 6). The first degree MB of *group* variable was selected from  $S^*$  (we denoted this CBN with 34 variables as  $S^*$ ): 3 genes that were direct causes (parents) of *group* (NFKB2, UBIAD1, and HEBP1), 12 genes that were direct effects (children) of *group* (FZD1, MAP2K2, BTNL8, TSPO, HOXB2, FOLH1, KIF11, SLC10A3, PSAT1, CACYBP, S100PBP, and ZFP36L2) and 20 genes that were direct causes (parents) of direct effects (children) of *group* (BTNL8, TEP1, L1TD1, TSPO, PSMB8, RPS6KB2, DLL3, TNFRSF11B, MAPK3, FASTKD3, ZNF273, PPP2CA, AGA, IFI30, ZDHHC6, INPP1, YWHAB, EPS8, RANBP6, and TATDN2).

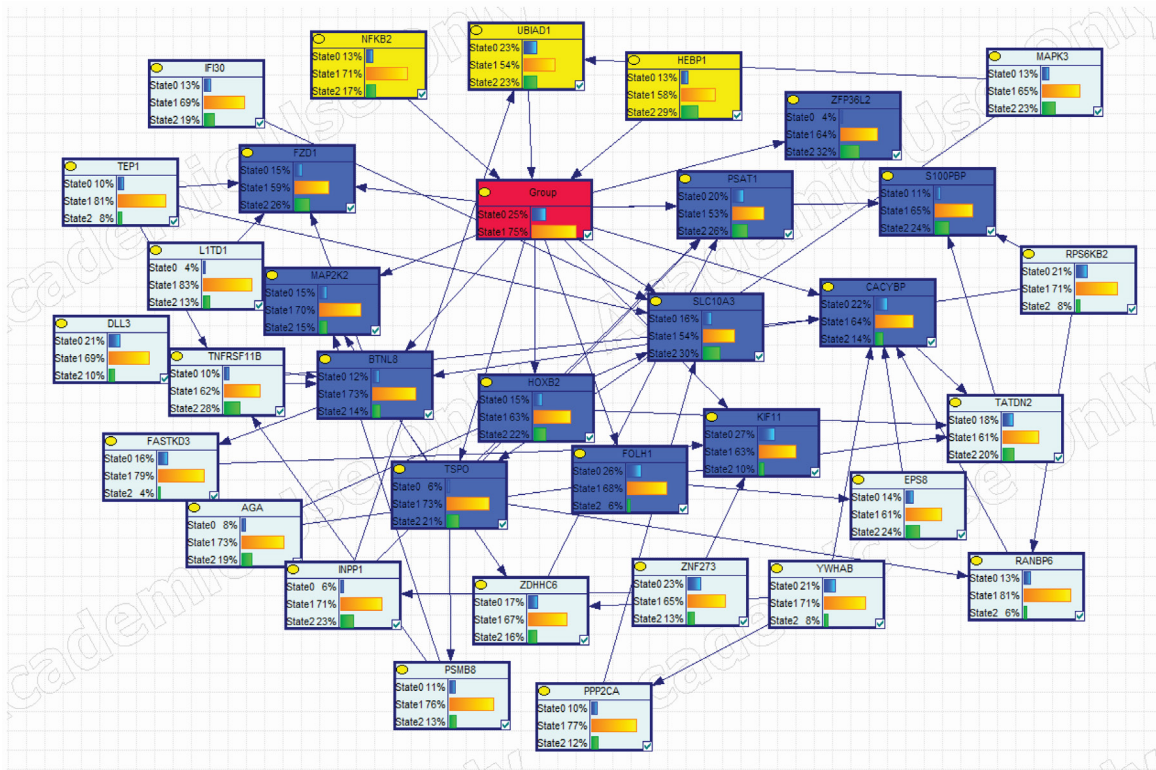
### 3. Learning causal Bayesian parameters

Probabilities (parameters) of the CBN with 34 variables that represents the first degree MB of *group* variable ( $S^*$ ) of CBN with BCRGs ( $S^*$ ) was learned from a new dataset (denoted as  $D^*$ ) with data of 34 variables extracted from data of 71 variables in  $D^*$  (48 subjects) (Fig. 7). The NFKB2 gene in  $S^*$  (direct cause [parent] of the *group* variable) was registered as BCRGs in the KEGG and NF- $\kappa$ B protein encoded by the NFKB2 gene is known as transcription factor promoting BC stem cell.[30] Among 12 direct effects (children) genes of BMBC (*group* variable) in  $S^*$ , 5 genes (SLC10A3, FZD1, BTNL8, KIF11, and CACYBP) were reported as the genes having close association with chemotherapy resistance.[31-35] In addition, KIF11, MAP2K2, PSAT1, TSPO, and ZFP36L2 genes within direct effects (children) of BMBC (*group* variable) were known as having relationship with cancer progression and metastasis.[36-39]

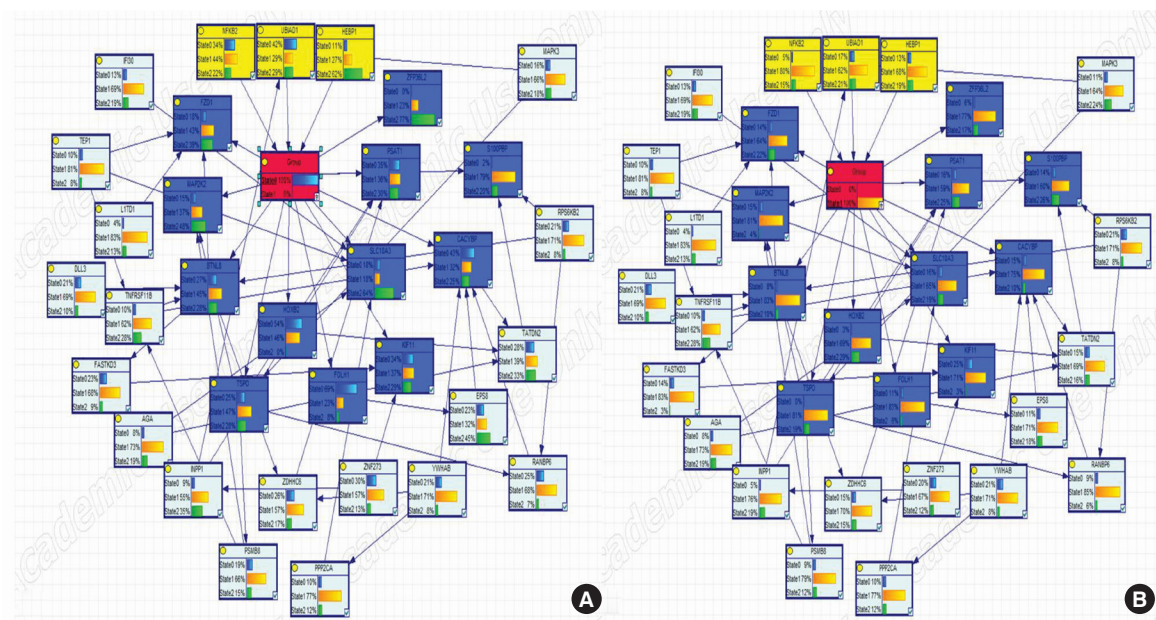
According to parameters learned in  $S^*$ , comparing a subject that has BMBC (*group* variable is in state 1, i.e., BM) with



**Fig. 6.** Causal Bayesian network with breast cancer relevant genes. A node filled with red color represented *group* (osteoblast or bone metastasis) node. The 13 nodes filled with green color represented breast cancer relevant genes.



**Fig. 7.** Causal Bayesian network structure using GeNIe. Picture show the causal Bayesian structure learned with existing data set using GeNIe. The parent, group, child, and co-parents nodes filled with yellow, red, blue, and bright blue, respectively. State 0 and 1 in group node represent probability of occurrence of osteoblast and bone metastasis, respectively. State 0, 1, and 2 in other nodes represent discretized value 0, 1, and 2, respectively.



**Fig. 8.** Causal Bayesian network structures according to different conditions of group node. (A, B) Pictures show changes of gene expressions dependent on 2 conditions as state 0 (probability of osteoblast occurrence)= 100% and state 1 (probability of bone metastasis occurrence)= 100%, respectively.

a subject that has osteoblast, among 15 genes that showed expression pattern change, FOLH1 and ZFP36L2 were the most significantly changed genes (FOLH1 expression changed from no change [state 1] to low [state 0] and ZFP36L2 expression changed from high [state 2] to no change [state 1]) (Fig. 8).

4. Assessment and validations

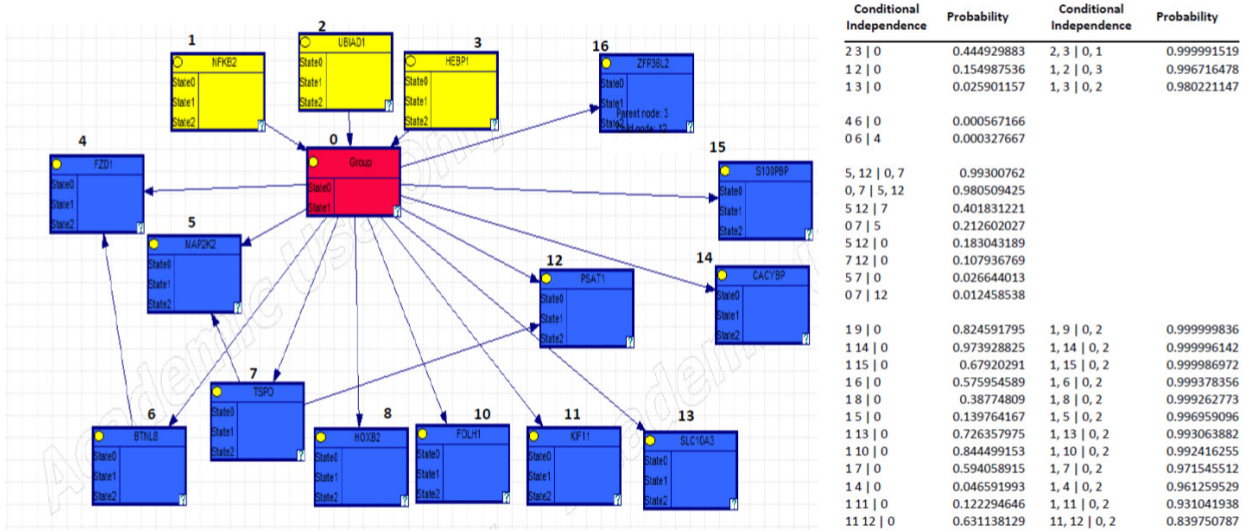
We further evaluated prediction performance of the CBN  $S^*$  parameterized by  $D^*$  with leave one out cross validation test achieved 68.38% (1,116 correct predictions out of  $34 \times 48$  cases). When we used only direct causes (parents) and effects (children) of BMBC (*group variable*) in  $S^*$  (i.e., 16 variables out of 34 variables) leave one out cross validation test retried 82.16% (631 correct predictions out of  $16 \times 48$  cases). AUROC predicting BMBC and osteoblast were 1 in  $S^*$  with the 16 variables (direct causes and effects of BMBC). Also in the CBN  $S^*$  parameterized by  $D^*$ , just using the 3 direct causes of BMBC (*group variable*) the probability of a subject having BM were over 90% (Table 2).

We have also calculated how likely any 2 variables are conditional independent (note that we are calculating the probabilities of conditional independencies, so the lower the probability is the higher the chances the variables are *not* independent) among the 16 variables (direct causes and effects of BMBC [*group variable*]) and *group variable*. [28] Calculated probabilities of conditional independence showed that conditional independency relationships (d-separation and d-connectivity [29]) among 16 variables and *group variable* in the  $S^*$  were agreeing with the structure of  $S^*$  (Fig. 9). These probabilities of conditional independence allowed us to further select 12 genes (Table 3) that showed higher relationships with BMBC (*group variable*). Figure 6 shows that except 1 gene, all 13 BCRGs are connected to BMBC (*group variable*) in CBN with BCRGs. Among the 12 genes, 6 genes (MAPK2, MAPK3, NFKB2, FZD1, DLL3, and RPS6KB2) were variables in  $S^*$ . We report 4 maximum RRs of combined gene expression patterns (Table 3). Table 3 shows UBIAD1, HEBP1, BTNL8, TSPO, PSAT1 and ZFP36L2 genes consistently appears in the highest RRs combined gene expression patterns. The difference of expression levels of ZFP36L2 was especially significant because its expression differed from low (represented as 0) to high (represented as 2) in all combination patterns whereas most of the other genes differed to only no change (rep-

Table 2. The prediction rates given different conditions

Cases of osteoblast and bone metastasis	Prediction rates given 3 parent genes information	Prediction rates given 3 parent and 12 child genes information	Prediction rates given 3 parent, 12 child, and 18 co-parent genes information
Osteoblast 1	0.125	6.35E-05	8.29929E-08
Osteoblast 2	0.0555556	1.22E-08	7.8259E-15
Osteoblast 3	0.125	8.01E-06	5.33354E-09
Osteoblast 4	0.0138889	1.45E-06	2.1411E-10
Osteoblast 5	0.34375	7.24E-06	6.81075E-12
Osteoblast 6	0.0555556	0.00358152	1.04724E-06
Osteoblast 7	0.125	0.008867	7.54625E-08
Osteoblast 8	0.0138889	0.0104353	0.00005475
Osteoblast 9	0.34375	0.00236364	1.13218E-05
Osteoblast 10	0.0138889	6.64E-06	1.65664E-08
Osteoblast 11	0.125	4.40E-06	1.16112E-08
Osteoblast 12	0.0138889	6.50E-08	2.09571E-11
Osteoblast 13	0.0138889	2.26E-08	2.32857E-10
Bone metastasis 1	0.98	0.998629	1
Bone metastasis 2	0.944444	0.998264	0.999999
Bone metastasis 3	0.986111	0.999468	1
Bone metastasis 4	0.944444	1	1
Bone metastasis 5	0.997778	1	1
Bone metastasis 6	0.944444	1	1
Bone metastasis 7	0.986111	0.999978	1
Bone metastasis 8	0.944444	0.999757	0.99999
Bone metastasis 9	0.997778	0.99999	0.999998
Bone metastasis 10	0.986111	1	1
Bone metastasis 11	0.98	0.999998	1
Bone metastasis 12	0.96875	1	1
Bone metastasis 13	0.997778	1	1
Bone metastasis 14	0.98	0.999993	1
Bone metastasis 15	0.997778	0.999999	1
Bone metastasis 16	0.997778	1	1
Bone metastasis 17	0.986111	0.999997	1
Bone metastasis 18	0.98	0.999988	1
Bone metastasis 19	0.997778	0.999974	0.999999
Bone metastasis 20	0.997778	0.999585	1
Bone metastasis 21	0.997778	0.999988	1
Bone metastasis 22	0.997778	0.99999	0.999995
Bone metastasis 23	0.944444	0.999435	1
Bone metastasis 24	0.986111	0.999997	1
Bone metastasis 25	0.997778	1	1
Bone metastasis 26	0.96875	0.994262	0.999555
Bone metastasis 27	0.875	0.996313	1
Bone metastasis 28	0.997778	1	1
Bone metastasis 29	0.997778	1	1
Bone metastasis 30	0.997778	0.999959	1
Bone metastasis 31	0.944444	0.999987	1
Bone metastasis 32	0.997778	0.999993	0.999999
Bone metastasis 33	0.875	0.999977	1
Bone metastasis 34	0.34375	0.999263	0.999964
Bone metastasis 35	0.96875	0.999941	1





**Fig. 9.** Probability of conditional independence among nodes. Left picture show the causal Bayesian network structure with 16 nodes. The parent, group, and co-parent nodes were filled with yellow, red, and blue colors, respectively. And right table shows the probabilities of conditional independences dependent on different conditions.

**Table 3.** Minimum combinations with maximum relative risks of combined gene expression patterns

Gene expression patterns	Probability	Relative risk
BMBC[1]   NFKB2[1], UBIAD1[1], HEBP1[1], FZD1[1], BTNL8[1], MAP2K2[0], TSPO[1], KIF11[1], SLC10A3[1], PSAT1[1], ZFP36L2[0]	1	74014862.80
BMBC[1]   NFKB2[1], UBIAD1[0], HEBP1[2], FZD1[1], BTNL8[2], MAP2K2[0], TSPO[0], KIF11[2], SLC10A3[2], PSAT1[0], ZFP36L2[2]	1.35108e-10	
BMBC[1]   NFKB2[1], UBIAD1[1], HEBP1[1], FZD1[1], BTNL8[1], MAP2K2[0], TSPO[1], KIF11[1], PSAT1[1], CACYBP[1], ZFP36L2[0]	1	58169972.66
BMBC[1]   NFKB2[1], UBIAD1[0], HEBP1[2], FZD1[1], BTNL8[2], MAP2K2[0], TSPO[2], KIF11[2], PSAT1[1], CACYBP[2], ZFP36L2[2]	1.7191e-8	
BMBC[1]   NFKB2[1], UBIAD1[1], HEBP1[1], FZD1[1], BTNL8[1], MAP2K2[0], TSPO[1], SLC10A3[1], PSAT1[1], CACYBP[1], ZFP36L2[0]	1	35717729.92
BMBC[1]   NFKB2[1], UBIAD1[0], HEBP1[2], FZD1[1], BTNL8[2], MAP2K2[0], TSPO[0], SLC10A3[2], PSAT1[1], CACYBP[2], ZFP36L2[2]	2.79973e-08	
BMBC[1]   NFKB2[1], UBIAD1[1], HEBP1[1], BTNL8 [1], MAP2K2[0], TSPO[1], KIF11[1], SLC10A3[1], PSAT1[1], CACYBP[1], ZFP36L2[0]	1	19295520.54
BMBC[1]   NFKB2[1], UBIAD1[0], HEBP1[2], BTNL8 [2], MAP2K2[0], TSPO[0], KIF11[2], SLC10A3[2], PSAT1[1], CACYBP[2], ZFP36L2[2]	5.18255e-08	

[0], low; [1], no change; [2], high.

resented as 1).

Searching through variable orders in the CBN (if variable A is (indirect) cause of variable B, then A is said to be in the higher order of B) has been suggested to be a different way of searching for best fitting CBN.[27] Using the 16 variables (direct causes and effects of BMBC [group variable]) the order with the best score was in the following order (higher to lower): NFKB2, HEBP1, group, FOLH1, ZFP36L2, HOXB2, PSAT1, TSPO, CACYBP, MAP2K2, UBIAD1, BTNL8, KIF11, S100BPB, and FZD1. Two direct cause (parent) genes (NFKB2

and HEBP1) of group variable in  $S^*$  (Fig. 6) consistently appeared in the higher order than group variable. However, UBIAD1 gene that was the direct cause (parent) of group variable in  $S^*$ , appeared in the lower order than group variable. The genes with the highest change in probability to be expressed high (State 2) given the fact that a subject has BMBC (FOLH1 and ZFP36L2), appeared right next to group variable in the order. The best CBN structure that was identified by the order analysis was significantly better in terms of log likelihood, however, the conditional inde-



dependencies of the 16 variables (direct causes and effects of BMBC [group variable]) were consistent with  $S^*$ .

## DISCUSSION

We have built a causal inference statistical model using a machine learning tool, i.e., CBN, from GEO gene expression data of osteoblast and BMBC. While building the causal inference statistical model, we have also incorporated medical and biological prior knowledge such as clinical medical practice and information of BCRGs registered in KEGG. We also validated the causal inference statistical model using data, statistical tools, and current knowledge.

Using cell line or animal model of BMBC, IL-6, TGF- $\beta$ , and TFF genes were over-expressed.[30,40-43] The recent article presented that 15 genes (APOPEC3B, ATL2, BBS1, C6orf61, C6orf167, MMS22L, KCNS1, MFAP3L, NIP7, NUP155, PALM2, PH-4, PGD5, SFT2D2, and STEAP3) were associated with the development of BMBC among patients.[43] Among 3 niches associated with BM, we considered the endosteal niche related with osteoblast as the critical key.[18] Therefore, we evaluated the causal network with the genes present in both BMBC and osteoblast, but excluded all of the genes present in the primary BC data set. Because of that, most signature genes, which reported in previous studies with BMBC, were not included in our study. We found 3 plausible upstream genes (NFKB2, UBIAD1, HEBP1) and 12 plausible downstream genes that had direct connection to BMBC. These 15 genes may be a good starting point to better understand the pathophysiology of BMBC. In turn, this will enable us to make efficient diagnosis and effective treatments of BMBC for patients who have primary BC without BM. Also, the plausible upstream genes could be powerful candidates for target therapy of BMBC. NFKB2, one of the plausible upstream genes, encode NF- $\kappa$ B protein that is a transcription factor known as a regulator of the immune system and promotes epithelial-to-mesenchymal transition and the metastasis of BC.[30,44,45] Although the protein encoded by UBIAD1 gene was reported as bladder tumor suppressor protein, there was no report about its role in tumor development or the progression of HEBP1.[46] The UBIAD1 gene had 3 direct cause (parent) genes in  $S^*$ , i.e., DVL1, INPP1, and MAPK3. DVL1 and INPP1 are registered as BCRGs in KEGG (Fig. 6). In addition, UBIAD1 gene is one of the genes that are reported in minimum combi-

nation of gene expression patterns that show maximum RR in this study. Therefore, although a study for the direct effect of HEBP1 gene on BMBC will be needed, however, it is plausible that the UBIAD1 may have an effect on BMBC through influences of other genes such as DVL1 and MAPK3.

We mentioned some genes related with chemotherapy resistance and cancer progression in the Results section. The CBN models that we learnt in this study can explain how the genes interact to develop BMBC (group variable that represent either a subject has osteoblast or BM), drug resistance, cancer progression and metastasis as published in earlier studies.[31-39] The role of those genes can be investigated as diagnostic candidates of BMBC in a future study. FOLH1 and ZFP36L2 had the remarkable changes of status in gene expression according to a subject with BMBC or osteoblast. Previous studies presented that the variants of proteins encoded by FOLH1 and ZFP36L2 genes were potential contributors of risk toward breast, pancreas and prostate cancers.[39,47] In the CBN structures presented in this study, the significant changes of FLOH1 AND ZFP36L2 according to state of group (BM or osteoblast) may suggest that the genes can be considered as candidates of diagnostic biomarker in BMBC. There were several limitations in the present study. Although the CBN in this study provided meaningful genes for BMBC and showed plausible causal structure among genes through machine learning techniques, the sample size were limited of only 48 human subjects and lacking the control subjects with neither of osteoblast or BM. Therefore, several signature genes related BM as CXCL12 and tumor necrosis factor-related apoptosis-inducing ligand, presented in the previous study using GSE 14017 and GSE 14018, were not selected in the final structure of the present study.[48] Despite having obtained excellent results on the validation tests, we should increase the number of human samples to get a point estimate closer to the mean value of the population. In parallel, we plan a future study to obtain orders of variables with probabilistic score using a new variable order search code that is currently being tested in Dr. Yoo's lab. Datasets used in this study that were downloaded from microarray GEO public repository had several limitations that are worth mentioning for proper interpretation of the results reported in this study: (1) there were gender ambiguity of donor osteoblast cell and normal bone tissue (thirteen subjects) making the result limited in generalizing among the wom-

en; (2) microarray data does not provide genomic sequences, resulting in unknown genomic variation information of subjects. Also, because the physiological osteoblasts and the osteoblasts in metastatic microenvironment are different, the responses in cancer metastasis may not be similar. Therefore, we should use the biologic data of the osteoblasts in metastatic site in the future study. However, we believe that the CBN in this study give us several leads for animal follow up studies or other human studies. Also we are planning to collect different types of gene expression data, e.g., RNA-sequencing, for validation and extension of the current causal inference statistical model that we have developed.

The CBN composed with 16 variables(3 direct causes and 12 direct effects of BMBC [group variable]) and seems to be a reasonable causal inference network with a high prediction rate. The direct cause and effect genes of BMBC may be useful candidates for early diagnosis and target therapy of BMBC. Among those genes, ZFP36L2 may be a meaningful candidate for a predictor of BMBC. In addition, CBN analysis may provide us with a more comprehensive understanding of the pathophysiology of BM.

## REFERENCES

- Randall RL. A promise to our patients with metastatic bone disease. *Ann Surg Oncol* 2014;21:4049-50.
- Lipton A, Theriault RL, Hortobagyi GN, et al. Pamidronate prevents skeletal complications and is effective palliative treatment in women with breast carcinoma and osteolytic bone metastases: long term follow-up of two randomized, placebo-controlled trials. *Cancer* 2000;88:1082-90.
- Mundy GR. Metastasis to bone: causes, consequences and therapeutic opportunities. *Nat Rev Cancer* 2002;2:584-93.
- Roodman GD. Mechanisms of bone metastasis. *N Engl J Med* 2004;350:1655-64.
- Hernandez RK, Adhia A, Wade SW, et al. Prevalence of bone metastases and bone-targeting agent use among solid tumor patients in the United States. *Clin Epidemiol* 2015; 7:335-45.
- Ren G, Esposito M, Kang Y. Bone metastasis and the metastatic niche. *J Mol Med (Berl)* 2015;93:1203-12.
- Fazilaty H, Mehdi-pour P. Genetics of breast cancer bone metastasis: a sequential multistep pattern. *Clin Exp Metastasis* 2014;31:595-612.
- Suva LJ, Washam C, Nicholas RW, et al. Bone metastasis: mechanisms and therapeutic opportunities. *Nat Rev Endocrinol* 2011;7:208-18.
- Sturge J, Caley MP, Waxman J. Bone metastasis in prostate cancer: emerging therapeutic strategies. *Nat Rev Clin Oncol* 2011;8:357-68.
- Coleman RE. Bone cancer in 2011: prevention and treatment of bone metastases. *Nat Rev Clin Oncol* 2011;9:76-8.
- Brook N, Brook E, Dharmarajan A, et al. Breast cancer bone metastases: pathogenesis and therapeutic targets. *Int J Biochem Cell Biol* 2018;96:63-78.
- Valastyan S, Weinberg RA. Tumor metastasis: molecular insights and evolving paradigms. *Cell* 2011;147:275-92.
- Deo RC. Machine learning in medicine. *Circulation* 2015; 132:1920-30.
- Deo RC, Nallamotheu BK. Learning about machine learning: the promise and pitfalls of big data and the electronic health record. *Circ Cardiovasc Qual Outcomes* 2016;9:618-20.
- Nemzek JA, Hodges AP, He Y. Bayesian network analysis of multi-compartmentalized immune responses in a murine model of sepsis and direct lung injury. *BMC Res Notes* 2015; 8:516.
- Yoo C, Ramirez L, Liuzzi J. Big data analysis using modern statistical and machine learning methods in medicine. *Int Neurolog J* 2014;18:50-7.
- Kang Y, Siegel PM, Shu W, et al. A multigenic program mediating breast cancer metastasis to bone. *Cancer Cell* 2003; 3:537-49.
- Ottewell PD. The role of osteoblasts in bone metastasis. *J Bone Oncol* 2016;5:124-7.
- Haider MT, Hohen I, Dear TN, et al. Modifying the osteoblastic niche with zoledronic acid in vivo-potential implications for breast cancer bone metastasis. *Bone* 2014;66:240-50.
- Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res* 2013;41:D991-5.
- Quackenbush J. Microarray data normalization and transformation. *Nat Genet* 2002;32 Suppl:496-501.
- Yu J, Smith VA, Wang PP, et al. Advances to Bayesian network inference for generating causal networks from observational biological data. *Bioinformatics* 2004;20:3594-603.
- Bielza C, Larrañaga P. Bayesian networks in neuroscience: a survey. *Front Comput Neurosci* 2014;8:131.

24. Adabor ES, Acquah-Mensah GK, Oduro FT. SAGA: a hybrid search algorithm for Bayesian Network structure learning of transcriptional regulatory networks. *J Biomed Inform* 2015;53:27-35.
25. Bindea G, Galon J, Mlecnik B. CluePedia Cytoscape plugin: pathway insights using integrated experimental and in silico data. *Bioinformatics* 2013;29:661-3.
26. Nishida K, Ono K, Kanaya S, et al. KEGGscape: a cytoscape app for pathway data integration. *F1000Res* 2014;3:144.
27. Agostinho NB, Machado KS, Werhli AV. Inference of regulatory networks with a convergence improved MCMC sampler. *BMC Bioinformatics* 2015;16:306.
28. Scutari M, Denis JB. *Bayesian networks: with examples in R*. Boca Raton, FL: CRC Press; 2014.
29. Charniak E. Bayesian networks without tears. *AI Mag* 1991; 12:50-63.
30. Kendellen MF, Bradford JW, Lawrence CL, et al. Canonical and non-canonical NF-kappaB signaling promotes breast cancer tumor-initiating cells. *Oncogene* 2014;33:1297-305.
31. Wang Q, Lu F, Lan R. RNA-sequencing dissects the transcriptome of polyploid cancer cells that are resistant to combined treatments of cisplatin with paclitaxel and docetaxel. *Mol Biosyst* 2017;13:2125-34.
32. Zhang H, Zhang X, Wu X, et al. Interference of Frizzled 1 (FZD1) reverses multidrug resistance in breast cancer cells through the Wnt/beta-catenin pathway. *Cancer Lett* 2012; 323:106-13.
33. Nakao T, Iwata T, Hotchi M, et al. Prediction of response to preoperative chemoradiotherapy and establishment of individualized therapy in advanced rectal cancer. *Oncol Rep* 2015;34:1961-7.
34. Jiang M, Zhuang H, Xia R, et al. KIF11 is required for proliferation and self-renewal of docetaxel resistant triple negative breast cancer cells. *Oncotarget* 2017;8:92106-18.
35. Shi Y, Hu W, Yin F, et al. Regulation of drug sensitivity of gastric cancer cells by human calcyclin-binding protein (CacyBP). *Gastric Cancer* 2004;7:160-6.
36. Huth HW, Albarnaz JD, Torres AA, et al. MEK2 controls the activation of MKK3/MKK6-p38 axis involved in the MDA-MB-231 breast cancer cell survival: correlation with cyclin D1 expression. *Cell Signal* 2016;28:1283-91.
37. Gao S, Ge A, Xu S, et al. PSAT1 is regulated by ATF4 and enhances cell proliferation via the GSK3beta/beta-catenin/cyclin D1 signaling pathway in ER-negative breast cancer. *J Exp Clin Cancer Res* 2017;36:179.
38. Mukherjee S, Das SK. Translocator protein (TSPO) in breast cancer. *Curr Mol Med* 2012;12:443-57.
39. Yonemori K, Seki N, Kurahara H, et al. ZFP36L2 promotes cancer cell aggressiveness and is regulated by antitumor microRNA-375 in pancreatic ductal adenocarcinoma. *Cancer Sci* 2017;108:124-35.
40. Rajski M, Vogel B, Baty F, et al. Global gene expression analysis of the interaction between cancer cells and osteoblasts to predict bone metastasis in breast cancer. *PLoS One* 2012; 7:e29743.
41. Mourskaia AA, Dong Z, Ng S, et al. Transforming growth factor-beta1 is the predominant isoform required for breast cancer cell outgrowth in bone. *Oncogene* 2009;28:1005-15.
42. Smid M, Wang Y, Zhang Y, et al. Subtypes of breast cancer show preferential site of relapse. *Cancer Res* 2008;68:3108-14.
43. Savci-Heijink CD, Halfwerk H, Koster J, et al. A novel gene expression signature for bone metastasis in breast carcinomas. *Breast Cancer Res Treat* 2016;156:249-59.
44. Yeo SK, French R, Spada F, et al. Opposing roles of Nfkb2 gene products p100 and p52 in the regulation of breast cancer stem cells. *Breast Cancer Res Treat* 2017;162:465-77.
45. Kim B, Kim HH, Lee ZH. Alpha-tocopheryl succinate inhibits osteolytic bone metastasis of breast cancer by suppressing migration of cancer cells and receptor activator of nuclear factor-kappaB ligand expression of osteoblasts. *J Bone Metab* 2018;25:23-33.
46. Fredericks WJ, McGarvey T, Wang H, et al. The bladder tumor suppressor protein TERE1 (UBIAD1) modulates cell cholesterol: implications for tumor progression. *DNA Cell Biol* 2011;30:851-64.
47. Naushad SM, Shree Divyya P, Janaki Ramaiah M, et al. Clinical utility of genetic variants of glutamate carboxypeptidase II in predicting breast cancer and prostate cancer risk. *Cancer Genet* 2015;208:552-8.
48. Zhang XH, Wang Q, Gerald W, et al. Latent bone metastasis in breast cancer tied to Src-dependent survival signals. *Cancer Cell* 2009;16:67-78.
49. Endo-Munoz L, Cumming A, Sommerville S, et al. Osteosarcoma is characterised by reduced expression of markers of osteoclastogenesis and antigen presentation compared with normal bone. *Br J Cancer* 2010;103:73-81.
50. Sadikovic B, Yoshimoto M, Chilton-MacNeill S, et al. Identifi-

- fication of interactive networks of gene expression associated with osteosarcoma oncogenesis by integrated molecular profiling. *Hum Mol Genet* 2009;18:1962-75.
51. Zhang XH, Wang Q, Gerald W, et al. Latent bone metastasis in breast cancer tied to Src-dependent survival signals. *Cancer Cell* 2009;16:67-78.
52. Xu J, Acharya S, Sahin O, et al. 14-3-3zeta turns TGF-beta's function from tumor suppressor to metastasis promoter in breast cancer by contextual changes of Smad partners from p53 to Gli2. *Cancer Cell* 2015;27:177-92.
53. Mourskaia AA, Amir E, Dong Z, et al. ABCC5 supports osteoclast formation and promotes breast cancer metastasis to bone. *Breast Cancer Res* 2012;14:R149.
54. Kimbung S, Kovacs A, Bendahl PO, et al. Claudin-2 is an independent negative prognostic factor in breast cancer and specifically predicts early liver recurrences. *Mol Oncol* 2014;8:119-28.