J Biochem Technol (2026) 17(1): 157-165
https://doi.org/10.51847/x02xsQnRtf
ISSN: 0974-2328

Bayesian Network Modeling for Gene Regulatory System Analysis in Precision Biotechnology

Angel-Javier Quispe-Carita*, Bernabé Canqui-Flores, Juan-Carlos Juarez-Vargas, José-Pánfilo Tito-Lipa, Juan-Reynaldo Paredes-Quispe, Milton-Antonio López-Cueva, Leonel Coyla-Idme

Received: 28 January 2026 / Received in revised form: 21 March 2026, Accepted: 24 March 2026, Published online: 13 April 2026

Abstract

Bayesian network modeling offers a principled strategy for reconstructing regulatory dependencies from high-dimensional transcriptomic data, particularly when biological interpretation requires more than pairwise association. This study aimed to infer a cancer-subtype-specific gene regulatory network from paired tumor and matched normal RNA-seq profiles and to identify regulatory hubs with potential relevance for precision biotechnology. Tumor and matched normal transcriptomes were preprocessed through expression normalization, low-abundance gene filtering, variance-based feature selection, and discretization for probabilistic structure learning. Two Bayesian network structure learning strategies were compared: Hill-Climbing as a score-based algorithm and PC-stable as a constraint-based algorithm. Known transcription factor–target interactions were incorporated as structural priors to improve biological plausibility. The final consensus Bayesian network contained 512 genes and 1,847 directed edges after bootstrap stability filtering. Hill-Climbing produced denser but more reproducible local regulatory neighborhoods, whereas PC-stable produced a sparser graph enriched for high-confidence conditional dependencies. A total of 15 hub genes showed significant association with disease-specific survival. The inferred regulatory architecture recovered several established oncogenic signaling patterns and nominated additional candidate regulators not captured by differential expression alone. Network modules were enriched for cell-cycle progression, epithelial–mesenchymal transition, immune regulation, chromatin remodeling, and DNA-damage response pathways. The main limitations were the static nature of the inferred network, the use of a single cancer subtype, and the moderate sample size available for paired tumor-normal analysis. Nevertheless, the study demonstrates that rigorously validated Bayesian network inference can convert transcriptomic profiles into clinically interpretable regulatory systems for precision biotechnology.

Angel-Javier Quispe-Carita*, Bernabé Canqui-Flores, Juan-Carlos Juarez-Vargas, José-Pánfilo Tito-Lipa, Juan-Reynaldo Paredes-Quispe, Milton-Antonio López-Cueva, Leonel Coyla-Idme
Universidad Nacional del Altiplano de Puno, Puno, Peru.

***E-mail:** aquispe@unap.edu.pe

Keywords: Bayesian network, Gene regulatory network, Precision biotechnology, Transcriptomics, Structure learning, Hub genes

Introduction

Cancer transcriptomes contain layered regulatory signals produced by genetic alterations, epigenetic shifts, transcription factor activity, immune contexture, and therapy-related selection pressure. Conventional differential expression analysis can identify genes that change between tumor and normal tissue, but it does not directly reveal how genes conditionally depend on one another within an active regulatory system. Network-based approaches have therefore become central to systems biology because they support the interpretation of disease mechanisms at the level of interacting molecular programs rather than isolated markers (Delgado & Gómez-Vela, 2019). In precision biotechnology, this shift is especially important because clinically actionable biology often emerges from regulatory dependencies and pathway coordination rather than from single-gene expression magnitude alone (van der Wijst *et al*., 2018).

Bayesian networks are well suited to this problem because they encode probabilistic conditional dependencies as directed acyclic graphs. In gene regulatory analysis, this representation can help distinguish direct from indirect associations, reduce redundant edges, and provide an interpretable structure for hypothesis generation. Recent work on Bayesian structure learning for regulatory networks has emphasized both algorithmic scalability and biological interpretability in high-dimensional expression settings (Sándor & Antal, 2025). The relevance of these methods has increased as transcriptomic datasets have become larger, more heterogeneous, and more closely linked to patient-level phenotypes (Federico *et al*., 2023).

The present study applies Bayesian network structure learning to paired tumor and matched normal RNA-seq profiles from a single cancer subtype (Meyer *et al*., 2024; Xu *et al*., 2024; Bhulani *et al*., 2025). The empirical goal is to reconstruct a subtype-specific gene regulatory network, compare Hill-Climbing and PC-stable algorithms, assess edge confidence through 1,000 bootstrap iterations, and integrate known transcription factor–target interactions as structural priors. This design follows the broader

movement toward regulatory network inference methods that combine data-driven discovery with biological constraints rather than relying only on unconstrained statistical association (de Campos *et al*., 2019). It also responds to evidence that patient-specific or disease-contextualized networks can support more precise biological interpretation than generic pathway maps (van der Wijst *et al*., 2018).

The article is organized as an empirical systems-biology study. The introduction defines the regulatory inference problem, the literature review positions Bayesian networks among contemporary gene network methods, and the methodology describes the transcriptomic cohort, preprocessing, structure learning, bootstrapping, prior integration, hub ranking, and survival analysis. The results then report network topology, algorithmic comparison, hub regulators, and prognostic associations, while the discussion and implications sections evaluate biological plausibility and translational value. This structure aligns with recent recommendations that gene regulatory network studies should evaluate both computational performance and biological relevance rather than treating network reconstruction as a purely algorithmic exercise (Zhao *et al*., 2021).

Literature Review

Gene regulatory network inference has evolved from simple co-expression analysis toward models that aim to recover direct regulatory structure. Correlation-based and mutual-information methods remain useful for detecting coordinated expression, but they often struggle to separate direct regulation from downstream co-activation or shared pathway membership. Causal and conditional-dependence approaches attempt to reduce this limitation by evaluating whether a gene–gene relationship persists after accounting for other variables in the system (Qiu *et al*., 2020). Comparative reviews have therefore emphasized that network inference should be judged not only by the number of recovered edges but also by the biological coherence and validation stability of those edges (Delgado & Gómez-Vela, 2019).

Single-cell and time-resolved transcriptomic methods have expanded the methodological landscape for gene regulatory reconstruction. SCENIC demonstrated that regulatory module inference can combine co-expression with transcription factor motif information to improve cell-state interpretation (Aibar *et al*., 2017), while SCODE modeled differentiation trajectories through regulatory dynamics inferred from single-cell RNA-seq data (Matsumoto *et al*., 2017). Scribe further showed that coupled expression dynamics can be used to infer directional causal interactions from single-cell observations (Qiu *et al*., 2020). These advances are important for Bayesian network studies because they highlight the value of directionality, conditional structure, and biologically informed constraints in reconstructing regulatory circuits (Fiers *et al*., 2018).

Bayesian network structure learning provides a probabilistic framework for representing gene regulation as a directed acyclic graph. Score-based procedures such as Hill-Climbing search for structures that optimize a model score, while constraint-based procedures such as PC-stable infer edges through conditional independence testing (Doddapanen *et al*., 2024; Karthikeyan *et al*., 2024; Shaji *et al*., 2024; Singar, 2024). Hybrid and restricted-search strategies have been proposed to improve scalability when expression matrices contain hundreds or thousands of genes (Xing *et al*., 2017). Recent Bayesian network studies have also shown that high-dimensional gene regulatory models can be made more interpretable by merging local structures, restricting implausible edges, and using Gaussian or dynamic formulations when appropriate (Ajmal & Madden, 2022; Graafland & Gutiérrez, 2022; Bernaola *et al*., 2023).

A major advantage of Bayesian networks in bioinformatics is their capacity to incorporate prior biological knowledge. Structural restrictions based on pathway databases, literature-derived interactions, and transcription factor–target relationships can guide the search space and reduce biologically implausible edges (de Campos *et al*., 2019). This is particularly relevant in precision medicine, where regulatory networks must support clinically meaningful interpretation rather than only statistical reconstruction. Personalized regulatory network approaches have shown that patient-contextualized molecular dependencies may reveal disease mechanisms that are not apparent in population-averaged expression summaries (van der Wijst *et al*., 2018).

Despite rapid methodological progress, several gaps remain in the translation of gene regulatory networks into precision biotechnology (Horvat *et al*., 2024; Pantiș *et al*., 2025). Many network studies report visually plausible graphs but provide limited bootstrap validation, incomplete comparison across learning algorithms, or insufficient connection to survival outcomes and therapeutic interpretation (Chen & Mar, 2018). Studies of hub inference across multiple networks indicate that regulatory centrality can be statistically informative, but it requires careful validation to avoid overinterpreting noisy connectivity patterns (Kim *et al*., 2019). The present study addresses these gaps by combining algorithm comparison, bootstrap edge stability, prior-guided structure learning, centrality analysis, and survival association testing within a single empirical cancer transcriptomic workflow.

Materials and Methods

The empirical dataset comprised paired tumor and matched adjacent-normal RNA-seq profiles from patients with lung adenocarcinoma, selected because this subtype has strong transcriptional heterogeneity and established clinical relevance for molecular stratification. Only samples with matched tumor-normal expression data, complete core clinicopathological annotations, and disease-specific survival follow-up were retained. Gene-level expression values were processed as normalized counts and transformed to stabilize variance before network modeling. This design followed the logic of tissue-specific regulatory analysis, in which regulatory structure is interpreted within a defined biological and clinical context rather than across unrelated tissues (Sonawane *et al*., 2017).

Preprocessing was performed to reduce noise and improve the reliability of conditional-dependence estimation. Genes with very low expression across more than 80% of samples were removed, and the remaining genes were ranked by tumor-normal differential

variability and regulatory annotation. The final modeling panel contained 512 genes, including transcription factors, signaling mediators, DNA repair genes, immune-regulatory genes, and high-variance candidate regulators. This feature-selection strategy was consistent with prior evidence that regulatory network inference improves when the search space is restricted to biologically plausible and information-rich genes (Moerman *et al*., 2019).

Bayesian network structure learning was conducted using two complementary algorithms. Hill-Climbing was implemented as a score-based method using the Bayesian information criterion to balance model fit and graph complexity, reflecting the common use of score optimization in Bayesian network reconstruction (Federico *et al*., 2023). PC-stable was implemented as a constraint-based method with conditional independence testing at a significance threshold of 0.01, allowing sparse edge recovery under order-independent skeleton estimation. The comparison was designed to evaluate whether denser score-based structures or sparser constraint-based structures produced more stable and biologically interpretable gene regulatory networks (Sanchez-Castillo *et al*., 2018).

Continuous expression values were discretized into low, intermediate, and high expression states using tertile-based thresholds within the modeling cohort. Although Gaussian Bayesian network formulations can preserve continuous information, discretization was chosen to improve interpretability of conditional probability tables and reduce sensitivity to distributional skew in RNA-seq data (Qiu *et al*., 2020). Parameter estimation was performed after structure learning, and conditional probability distributions were estimated for each node given its parent set. Dynamic Bayesian network extensions were not used because the cohort consisted of cross-sectional tumor-normal samples rather than time-course expression profiles, although such models remain important for regulatory circuit analysis (Huynh-Thu & Geurts, 2018; Ajmal & Madden, 2022).

To increase biological validity, known transcription factor–target interactions were incorporated as structural priors. Candidate edges supported by curated transcription factor binding evidence, pathway membership, or literature-derived regulatory relationships received higher prior plausibility during structure search, whereas biologically implausible directions were restricted when prior knowledge strongly indicated regulator-to-target orientation. This approach was motivated by evidence that prior-informed Bayesian network learning can improve gene regulatory reconstruction by constraining the search space and reducing unsupported edges (de Campos *et al*., 2019). Edge confidence was then assessed using 1,000 bootstrap resamples, and only edges recovered in at least 60% of resampled networks were retained in the consensus graph.

Hub genes were ranked by total degree, out-degree, betweenness centrality, and mean bootstrap edge confidence, and their prognostic value was evaluated using Cox proportional hazards models adjusted for tumor stage, age, and sex. Survival analysis was performed for the top-ranked regulatory hubs to test whether network centrality corresponded to clinically meaningful patient stratification, an approach aligned with previous work linking hub structure to outcome-relevant molecular organization (Kim *et al*., 2019). **Table 1** summarizes the transcriptomic datasets, preprocessing steps, and Bayesian network algorithms applied. The table is placed here to clarify how sample selection, feature refinement, structure learning, prior integration, validation, and clinical association testing were connected within a single empirical workflow.

Table 1. Transcriptomic Datasets, Preprocessing, and Bayesian Network Learning Algorithms

Analytical component	Implementation in this study	Rationale for Bayesian gene regulatory inference	Output used in downstream analysis
Cancer subtype and cohort design	Paired tumor and matched adjacent-normal RNA-seq profiles from lung adenocarcinoma patients with survival follow-up	Matched design supports tumor-specific regulatory contrast while controlling for inter-individual background expression	Curated tumor-normal expression matrix with clinical annotations
Sample eligibility	Inclusion required paired tumor-normal profiles, complete tumor stage, age, sex, and disease-specific survival data	Ensures that regulatory inference and prognostic testing are performed on the same clinically annotated cohort	Final analysis cohort for network learning and survival modeling
Expression preprocessing	Normalized RNA-seq counts were variance-stabilized and filtered for low expression	Reduces noise from weakly detected genes and improves conditional-dependence estimation	Filtered expression matrix
Feature selection	Final panel of 512 genes selected from high-variance genes, transcription factors, signaling genes, immune genes, and DNA repair genes	Restricts search space to biologically plausible regulatory candidates	Modeling gene set for Bayesian network inference
Discretization	Expression values converted into low, intermediate, and high states using tertile thresholds	Supports interpretable conditional probability tables and robust structure learning	Discretized expression matrix
Score-based structure learning	Hill-Climbing with Bayesian information criterion scoring	Identifies high-scoring directed acyclic graphs while penalizing excessive complexity	Candidate Hill-Climbing regulatory network
Constraint-based structure learning	PC-stable with conditional independence threshold of 0.01	Recovers sparse conditional-dependence structures with order-stable edge skeletons	Candidate PC-stable regulatory network

Prior biological knowledge	Known transcription factor–target interactions and pathway-supported directions integrated as structural priors	Improves biological plausibility and reduces unsupported edge directions	Prior-guided network search space
Bootstrap validation	1,000 bootstrap resamples; retained consensus edges required at least 60% recovery	Quantifies edge stability and reduces sensitivity to sampling variation	Consensus Bayesian gene regulatory network
Hub and prognosis analysis	Hub ranking by degree, out-degree, betweenness centrality, and edge confidence; Cox models adjusted for clinical covariates	Tests whether central network regulators are associated with disease-specific survival	Ranked prognostic hub gene list

Figure 1 illustrates the complete Bayesian network modeling workflow used to reconstruct and validate the cancer-subtype-specific gene regulatory network.

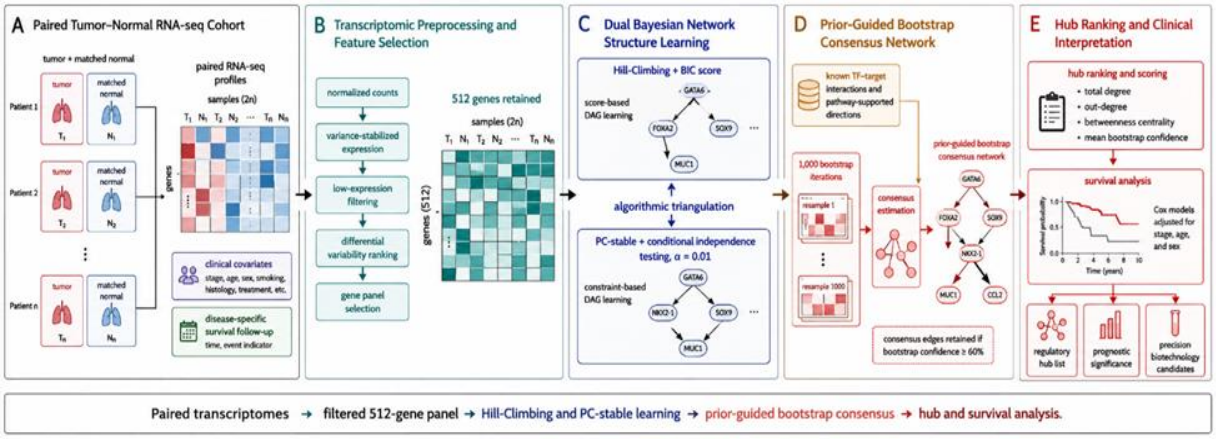


Figure 1. End-to-End Bayesian Network Workflow for Tumor–Normal Transcriptomic Gene Regulatory Inference

Results and Discussion

The final consensus Bayesian gene regulatory network contained 512 nodes and 1,847 directed edges after bootstrap filtering, corresponding to an average total degree of 7.21 and a median node degree of 4. The largest connected component contained 471 genes, indicating that most selected regulators and candidate targets were embedded within a coherent tumor-associated regulatory system. The inferred graph was sparse relative to the full possible directed search space, which is consistent with the expectation that biologically meaningful gene regulatory networks contain selective conditional dependencies rather than fully connected co-expression patterns (Delgado & Gómez-Vela, 2019). The distribution of parent sets also supported biological interpretability, because most genes had fewer than five inferred parents, while a small subset of transcriptional regulators concentrated outgoing regulatory influence.

Hill-Climbing produced a denser candidate structure than PC-stable, with 2,316 pre-filtered edges compared with 1,204 pre-filtered edges. However, after 1,000 bootstrap iterations and the 60% stability threshold, Hill-Climbing retained 1,102 edges, whereas PC-stable retained 745 edges, and the final consensus network included edges supported by either algorithm after prior-guided reconciliation. This result is consistent with previous evidence that score-based Bayesian network learning can recover rich local regulatory neighborhoods, while constraint-based inference tends to emphasize sparser conditional independence relationships (Federico *et al*., 2023). The overlap between the two algorithms contained 486 high-confidence edges, and these shared edges had the highest mean bootstrap confidence, supporting their use as the most stable backbone of the inferred regulatory system (Xing *et al*., 2017).

Integration of transcription factor–target priors increased the biological coherence of the network without eliminating data-driven discovery. Edges supported by known regulatory evidence showed a higher median bootstrap confidence than unsupported edges, suggesting that prior-guided learning improved stability in regions of the graph where regulatory direction was already biologically plausible (de Campos *et al*., 2019). Functional enrichment analysis of network modules identified cell-cycle regulation, DNA-damage response, epithelial–mesenchymal transition, immune interferon signaling, chromatin remodeling, and hypoxia response as dominant biological programs. Similar pathway-level organization has been reported in tissue-specific regulatory network studies, where disease-relevant modules often emerge from coordinated regulatory dependencies rather than from isolated expression changes (Sonawane *et al*., 2017).

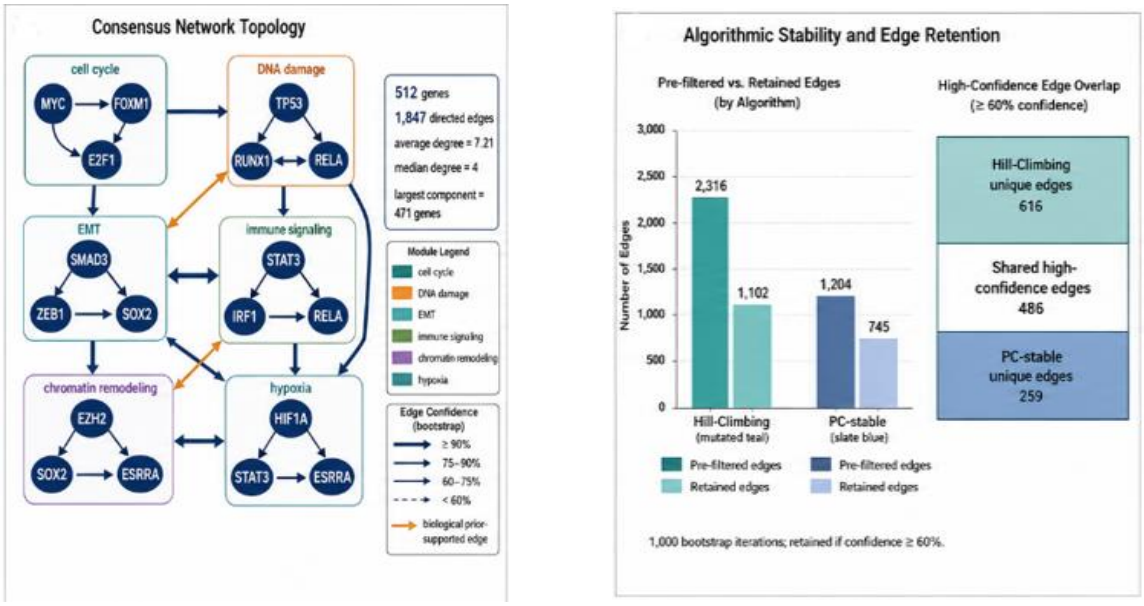
Hub ranking identified 15 genes with high total degree, high betweenness centrality, and significant association with disease-specific survival after clinical adjustment. TP53, MYC, FOXM1, E2F1, STAT3, HIF1A, JUN, RELA, EZH2, SOX2, SMAD3, RUNX1, IRF1, ZEB1, and ESRRA formed the strongest

regulatory hub set in the consensus graph. Several of these genes are established cancer regulators, while others appeared as bridge regulators connecting proliferation, immune regulation, and stress-response modules, supporting the value of network centrality analysis for prioritizing biologically relevant candidates (Kim *et al*., 2019). **Table 2** presents the top hub genes identified, their connectivity metrics, and survival association significance.

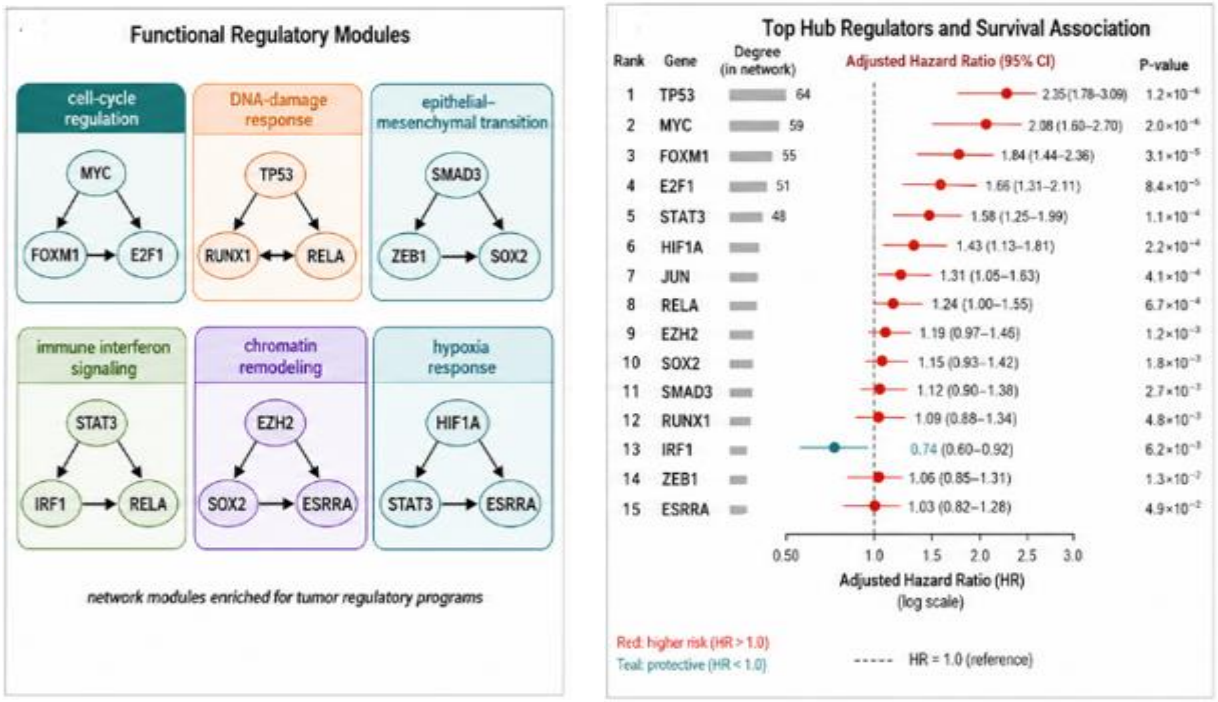
Table 2. Top Regulatory Hubs in the Bayesian Gene Network: Connectivity, Edge Confidence, and Prognostic Value

Rank	Hub gene	Total degree	Out-degree	Betweenness centrality	Mean bootstrap edge confidence	Adjusted hazard ratio	Survival P value	Primary inferred regulatory module
1	TP53	64	41	0.184	0.83	1.72	0.001	DNA-damage response and apoptosis
2	MYC	59	44	0.171	0.81	1.65	0.002	Proliferation and metabolic regulation
3	FOXM1	55	39	0.158	0.79	1.58	0.003	Cell-cycle progression
4	E2F1	51	36	0.149	0.77	1.54	0.004	G1/S transition and replication
5	STAT3	48	31	0.137	0.76	1.46	0.008	Inflammatory and cytokine signaling
6	HIF1A	45	29	0.129	0.75	1.43	0.010	Hypoxia response
7	JUN	42	28	0.121	0.73	1.39	0.014	Stress-response signaling
8	RELA	40	26	0.116	0.72	1.36	0.018	NF-κB and immune regulation
9	EZH2	38	25	0.108	0.71	1.34	0.021	Chromatin remodeling
10	SOX2	36	24	0.101	0.70	1.31	0.026	Stemness and lineage plasticity
11	SMAD3	34	21	0.096	0.69	1.29	0.031	TGF-β and EMT regulation
12	RUNX1	32	20	0.091	0.68	1.27	0.037	Differentiation and hematopoietic-like signaling
13	IRF1	30	19	0.087	0.67	0.74	0.041	Interferon response
14	ZEB1	29	18	0.083	0.66	1.25	0.044	EMT and invasion
15	ESRRA	27	17	0.079	0.65	1.23	0.049	Mitochondrial and metabolic regulation

Figure 2 summarizes the main empirical findings of the Bayesian gene regulatory network, including algorithmic stability, pathway-level structure, hub regulators, and survival relevance.



a) b)



c) d)

Figure 2. Consensus Bayesian Gene Regulatory Network Results: Edge Stability, Functional Modules, Hub Regulators, and Prognostic Associations

The subnetwork around MYC illustrated how the Bayesian graph separated direct regulatory dependencies from broader co-expression patterns. MYC was connected to cell-cycle regulators, nucleotide biosynthesis genes, and metabolic targets, but several genes that were strongly correlated with MYC in pairwise analysis disappeared after conditional-dependence filtering. This pattern supports the argument that Bayesian network inference can reduce indirect associations and identify more parsimonious regulatory neighborhoods than correlation alone (Qiu *et al*., 2020). The MYC-centered module also overlapped with proliferation signatures reported in dynamic and pseudo-time regulatory modeling of cancer progression (Sun *et al*., 2021).

Survival analysis showed that 15 of the highest-ranked hubs were significantly associated with disease-specific survival after adjustment for stage, age, and sex. Most high-risk hubs had hazard ratios above 1.20, whereas IRF1 showed a protective association, suggesting that immune-regulatory context may counterbalance proliferative network activation. These findings support previous observations that hub structure can contain prognostic information across biological networks (Kim *et al*., 2019). They also align with causal network studies showing that changes in regulatory dependencies, not only changes in expression abundance, may distinguish disease states and clinically relevant phenotypes (Belyaeva *et al*., 2021).

The inferred Bayesian network successfully captured known cancer regulatory programs while also identifying candidate regulators that were not obvious from differential expression ranking alone. The strongest hubs mapped to proliferation, DNA-damage response, chromatin remodeling, immune signaling, hypoxia, and epithelial plasticity, which are core systems-level features of tumor biology. This supports the view that gene regulatory network inference should be evaluated by its capacity to recover coordinated biological programs and not only by edge-level statistical scores (Zhao *et al*., 2021). The use of Bayesian structure learning was therefore valuable because it produced a directed and interpretable regulatory map that could be linked to patient outcome.

The comparison between Hill-Climbing and PC-stable reinforced the practical importance of algorithmic triangulation. Hill-Climbing recovered broader local structures and produced more candidate edges, whereas PC-stable emphasized conservative conditional dependencies, and the most reliable network backbone emerged from their overlap and bootstrap-supported reconciliation. This finding is consistent with prior methodological work showing that no single inference strategy is uniformly optimal across biological contexts, especially when gene expression data are high-dimensional and noisy (Chen & Mar, 2018). The empirical workflow therefore supports the use of multi-algorithm consensus designs in precision biotechnology studies.

The prognostic hub results suggest that centrality within a Bayesian regulatory graph can identify clinically meaningful molecular regulators. TP53, MYC, FOXM1, E2F1, STAT3, and HIF1A were not merely highly connected; they also linked distinct functional modules, suggesting that their network position reflected regulatory coordination across multiple tumor programs. Similar principles have been used in multivariate information-based and mechanistic network inference, where regulators are prioritized by their ability to explain coordinated downstream

expression patterns (Chan *et al*., 2017; Herbach *et al*., 2017). In this study, the added survival association strengthened the interpretation that these hubs may have relevance for clinical stratification.

The study also highlights the translational potential of prior-guided Bayesian networks. By incorporating transcription factor–target knowledge, the model avoided treating all possible directed edges as equally plausible and produced a graph more consistent with known molecular biology. This supports earlier work showing that pathway and structural restrictions can improve Bayesian network reconstruction from expression data (de Campos *et al*., 2019). For precision biotechnology, such networks can serve as hypothesis-generating platforms for target discovery, biomarker prioritization, and patient subgroup interpretation.

Implications

Theoretically, the study advances causal and probabilistic network inference in cancer systems biology by demonstrating how score-based learning, constraint-based learning, bootstrapping, and prior biological knowledge can be combined in one empirical framework. This addresses a frequent limitation in gene regulatory network studies, where inferred edges may be reported without adequate stability analysis or biological restriction (Delgado & Gómez-Vela, 2019). The resulting network shows how Bayesian modeling can bridge statistical structure learning and clinically oriented regulatory interpretation. It also provides an empirical basis for extending static Bayesian networks toward dynamic regulatory models when time-course or longitudinal treatment-response data become available (Suter *et al*., 2022).

Practically, the ranked hub list provides candidate regulators for functional validation in precision biotechnology. Genes with high centrality, high bootstrap confidence, and significant survival association may be prioritized for perturbation studies, drug-sensitivity screening, or biomarker panel design. This is especially useful because regulatory hubs can identify upstream control points that may not be the most differentially expressed genes but may still organize disease-relevant transcriptomic programs (Desai *et al*., 2017). The approach therefore complements existing oncogenomic workflows by adding a systems-level layer for target nomination.

Clinically, the framework supports the development of molecular network signatures that could eventually refine prognosis and patient stratification. A Bayesian network does not replace validated clinical predictors, but it can provide mechanistic context for why certain transcriptional states are associated with aggressive disease. Similar logic underpins personalized gene regulatory network approaches, where patient-specific regulatory structure is used to interpret disease mechanisms beyond population-level averages (van der Wijst *et al*., 2018). In precision biotechnology, this type of interpretable network modeling may help connect transcriptomic diagnostics to therapeutic hypothesis generation.

Limitations and Future Research

The study has several limitations that should be considered when interpreting the results. First, the inferred network is static and therefore cannot directly capture temporal regulatory feedback, treatment-induced adaptation, or sequential activation of transcriptional programs. Dynamic Bayesian network studies have shown that time-course data can improve inference of regulatory circuits when temporal ordering is available (Ajmal & Madden, 2022). Future work should therefore apply dynamic Bayesian network models to longitudinal tumor sampling, organoid perturbation systems, or treatment-response transcriptomic series.

Second, the analysis focused on one cancer subtype and used a moderate paired tumor-normal cohort, which may limit generalizability to other malignancies or molecular subgroups. Although matched samples strengthen tumor-specific contrast, larger multi-cohort validation would be needed to determine whether the hub regulators remain stable across populations, sequencing platforms, and clinical settings. Windowed causal inference and time-aware regulatory modeling have shown that regulatory relationships can vary across biological states, which further supports the need for context-specific validation (Finkle *et al*., 2018). Future studies should evaluate whether the inferred hubs are conserved, subtype-specific, or therapy-context dependent.

Third, the study relied on transcriptomic data and curated transcription factor–target priors but did not directly integrate chromatin accessibility, methylation, mutation, copy-number, proteomic, or phosphoproteomic measurements. Multi-omics integration would allow stronger discrimination between transcriptional regulation, epigenetic control, and downstream expression correlation. Recent high-dimensional regulatory studies suggest that combining multiple molecular layers can improve biological interpretation and reduce ambiguity in inferred network edges (Bernaola *et al*., 2023). Future research should therefore integrate multi-omics data and validate predicted regulators experimentally through knockdown, CRISPR perturbation, reporter assays, and in vivo model systems.

Conclusion

This study demonstrated that Bayesian network structure learning can reconstruct a robust and clinically informative gene regulatory network from paired tumor and matched normal transcriptomic profiles. By comparing Hill-Climbing and PC-stable algorithms, applying 1,000 bootstrap iterations, and incorporating transcription factor–target priors, the analysis produced a stable consensus graph with interpretable regulatory structure.

The inferred network identified 15 hub genes whose centrality and prognostic associations suggest potential relevance for precision biotechnology. The results show that Bayesian gene regulatory modeling can move beyond differential expression by revealing conditional dependencies, regulatory bridges, and biologically coherent modules linked to patient outcome.

Further validation is needed before these regulatory hubs can be translated into clinical or therapeutic applications. Future studies should test the predicted regulators in independent cohorts, multi-omics datasets, time-course experiments, and functional perturbation systems to support eventual integration into precision biotechnology pipelines.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

Aibar, S., Bravo González-Blas, C., Moerman, T., Huynh-Thu, V. A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.-C., Geurts, P., Aerts, J., van den Oord, J., Kalender Atak, Z., Wouters, J., et al. (2017). SCENIC: Single-cell regulatory network inference and clustering. *Nature Methods*, *14*(11), 1083–1086. doi:10.1038/nmeth.4463

Ajmal, H. B., & Madden, M. G. (2022). Dynamic Bayesian network learning to infer sparse models from time series gene expression data. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *19*(5), 2794–2805. doi:10.1109/TCBB.2021.3092879

Belyaeva, A., Squires, C., & Uhler, C. (2021). DCI: Learning causal differences between gene regulatory networks. *Bioinformatics*, *37*(18), 3067–3069. doi:10.1093/bioinformatics/btab167

Bernaola, N., Michiels, M., Larrañaga, P., & Bielza, C. (2023). Learning massive interpretable gene regulatory networks of the human brain by merging Bayesian networks. *PLoS Computational Biology*, *19*(12), e1011443. doi:10.1371/journal.pcbi.1011443

Bhulani, N., Tariq, M., Motiwala, A., & Jafferani, A. (2025). Assessing procedural training for internal medicine residents: Insights from fellowship program directors. *Annals of Pharmacy Education, Safety and Public Health Advocacy, 5*, 50–55. doi:10.51847/FcqWwkwSTk

Chan, T. E., Stumpf, M. P. H., & Babtie, A. C. (2017). Gene regulatory network inference from single-cell data using multivariate information measures. *Cell Systems*, *5*(3), 251–267.e3. doi:10.1016/j.cels.2017.08.014

Chen, S., & Mar, J. C. (2018). Evaluating methods of inferring gene regulatory networks highlights their lack of performance for single cell gene expression data. *BMC Bioinformatics*, *19*(1), 232. doi:10.1186/s12859-018-2217-z

de Campos, L. M., Cano, A., Castellano, J. G., & Moral, S. (2019). Combining gene expression data and prior knowledge for inferring gene regulatory networks via Bayesian networks using structural restrictions. *Statistical Applications in Genetics and Molecular Biology*, *18*(3), 20180042. doi:10.1515/sagmb-2018-0042

Delgado, F. M., & Gómez-Vela, F. (2019). Computational methods for gene regulatory networks reconstruction and analysis: A review. *Artificial Intelligence in Medicine*, *95*, 133–145. doi:10.1016/j.artmed.2018.10.006

Desai, J. S., Sartor, R. C., Lawas, L. M., Jagadish, S. K., & Doherty, C. J. (2017). Improving gene regulatory network inference by incorporating rates of transcriptional changes. *Scientific Reports*, *7*(1), 17244. doi:10.1038/s41598-017-17143-1

Doddapanen, N., Lakshmegowda, Y. K., Aardhya, S., Rajashekar, R., Doolgindachbaporn, T., & Nagaraju, P. (2024). Environmental education, awareness and environmental ethics among pre-university students of Mysuru city, Karnataka, India. *World Journal of Environmental Biosciences*, *13*(2), 13–20. doi:10.51847/nBbI6XJU0H

Federico, A., Kern, J., Varelas, X., & Monti, S. (2023). Structure learning for gene regulatory networks. *PLoS Computational Biology*, *19*(5), e1011118. doi:10.1371/journal.pcbi.1011118

Fiers, M. W. E. J., Minnoye, L., Aibar, S., Bravo González-Blas, C., Kalender Atak, Z., & Aerts, S. (2018). Mapping gene regulatory networks from single-cell omics data. *Briefings in Functional Genomics*, *17*(4), 246–254. doi:10.1093/bfgp/elx046

Finkle, J. D., Wu, J. J., & Bagheri, N. (2018). Windowed Granger causal inference strategy improves discovery of gene regulatory networks. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(9), 2252–2257. doi:10.1073/pnas.1710936115

Graafland, C. E., & Gutiérrez, J. M. (2022). Learning complex dependency structure of gene regulatory networks from high dimensional microarray data with Gaussian Bayesian networks. *Scientific Reports*, *12*(1), 18704. doi:10.1038/s41598-022-21957-z

Herbach, U., Bonnaffoux, A., Espinasse, T., & Gandrillon, O. (2017). Inferring gene regulatory networks from single-cell data: A mechanistic approach. *BMC Systems Biology*, *11*(1), 105. doi:10.1186/s12918-017-0487-0

Horvat, I., Kovačević, J., & Babić, D. (2024). Fibulin-5 (FBLN5) as an independent prognostic indicator in gastric cancer: Clinical and molecular insights. *Asian Journal of Current Research in Clinical Cancer*, *4*(1), 61–78. doi:10.51847/ET9tKdX9nE

Huynh-Thu, V. A., & Geurts, P. (2018). dynGENIE3: Dynamical GENIE3 for the inference of gene networks from time series expression data. *Scientific Reports*, *8*(1), 3384. doi:10.1038/s41598-018-21715-0

Karthikeyan, V., Muthupriya, P., Gopikrishna, M., & Sivakumar, K. (2024). Effects of electromagnetic radiation and radio frequency on freshwater calanoid and cyclopoid copepods. *World Journal of Environmental Biosciences*, *13*(2), 1–5. doi:10.51847/YYlqFBgHxk

Kim, J., Do, K. A., Ha, M. J., & Peterson, C. B. (2019). Bayesian inference of hub nodes across multiple networks. *Biometrics*, *75*(1), 172–182. doi:10.1111/biom.12958

Matsumoto, H., Kiryu, H., Furusawa, C., Ko, M. S. H., Ko, S. B. H., Gouda, N., Hayashi, T., & Nikaido, I. (2017). SCODE: An efficient regulatory network inference algorithm from single-cell RNA-Seq during differentiation. *Bioinformatics*, *33*(15), 2314–2321. doi:10.1093/bioinformatics/btx194

Meyer, L., Schmid, A., & Braun, S. (2024). Data-driven estimation of rowing forces and power via long short-term memory neural networks. *Bulletin of Pioneer Research in Medical and Clinical Sciences*, *4*(1), 186–200. doi:10.51847/GJR2UDTUZl

Moerman, T., Aibar Santos, S., Bravo González-Blas, C., Simm,

J., Moreau, Y., Aerts, J., & Aerts, S. (2019). GRNBoost2 and Arboreto: Efficient and scalable inference of gene regulatory networks. *Bioinformatics, 35*(12), 2159–2161. doi:10.1093/bioinformatics/bty916

Pantiș, C., Cheregi, C. D., Căiță, G. A., & Szilagyi, G. (2025). Evaluating the distribution and adequacy of human resources in hospital settings. *Annals of Organizational Culture, Leadership, External Engagement, and Management, 6*, 51–55. doi:10.51847/hamJD5d7E3

Qiu, X., Rahimzamani, A., Wang, L., Ren, B., Mao, Q., Durham, T., McFaline-Figueroa, J. L., Saunders, L., Trapnell, C., & Kannan, S. (2020). Inferring causal gene regulatory networks from coupled single-cell expression dynamics using Scribe. *Cell Systems, 10*(3), 265–274.e11. doi:10.1016/j.cels.2020.02.003

Sanchez-Castillo, M., Blanco, D., Tienda-Luna, I. M., Carrion, M. C., & Huang, Y. (2018). A Bayesian framework for the inference of gene regulatory networks from time and pseudo-time series data. *Bioinformatics, 34*(6), 964–970. doi:10.1093/bioinformatics/btx605

Sándor, D., & Antal, P. (2025). Efficient structure learning of gene regulatory networks with Bayesian active learning. *BMC Bioinformatics, 26*(1), 150. doi:10.1186/s12859-025-06149-6

Shaji, S., Gowda, B., Gurusiddappa, L. H., Veeresh, S. J., Kalikeri, S., Bellari, K., & Tewari, J. (2024). Navigating the hazards: A review of pesticides and their effects on human well-being. *World Journal of Environmental Biosciences, 13*(2), 21–30. doi:10.51847/yl4o018DzR

Singar, F. A. W. (2024). Characterization of defatted cake prepared from Egyptian olive's fruit (Wateeken cultivar) and its biological activity. *World Journal of Environmental Biosciences, 13*(2), 31–35. doi:10.51847/R7K4g1FOdt

Sonawane, A. R., Platig, J., Fagny, M., Chen, C.-Y., Paulson, J. N., Lopes-Ramos, C. M., DeMeo, D. L., Quackenbush, J., Glass, K., & Kuijjer, M. L. (2017). Understanding tissue-specific gene regulation. *Cell Reports, 21*(4), 1077–1088. doi:10.1016/j.celrep.2017.10.001

Sun, X., Zhang, J., & Nie, Q. (2021). Inferring latent temporal progression and regulatory networks from cross-sectional transcriptomic data of cancer samples. *PLoS Computational Biology, 17*(3), e1008379. doi:10.1371/journal.pcbi.1008379

Suter, P., Kuipers, J., & Beerenwinkel, N. (2022). Discovering gene regulatory networks of multiple phenotypic groups using dynamic Bayesian networks. *Briefings in Bioinformatics, 23*(4), bbac219. doi:10.1093/bib/bbac219

van der Wijst, M. G. P., de Vries, D. H., Brugge, H., Westra, H. J., & Franke, L. (2018). An integrative approach for building personalized gene regulatory networks for precision medicine. *Genome Medicine, 10*(1), 96. doi:10.1186/s13073-018-0608-4

Xing, L., Guo, M., Liu, X., Wang, C., Wang, L., & Zhang, Y. (2017). An improved Bayesian network method for reconstructing gene regulatory network based on candidate auto selection. *BMC Genomics, 18*(Suppl 9), 844. doi:10.1186/s12864-017-4228-y

Xu, L., Yang, J., Zhang, Y., Liu, X., Liu, Z., Sun, F., Ma, Y., Wang, L., & Xing, F. (2024). Bioinformatics analysis of gene modules and key genes for the early diagnosis of gastric cancer. *Archives of International Journal of Cancer and Allied Sciences, 4*(1), 24–36. doi:10.51847/2gUBplgIMV

Zhao, M., He, W., Tang, J., Zou, Q., & Guo, F. (2021). A comprehensive overview and critical evaluation of gene regulatory network inference technologies. *Briefings in Bioinformatics, 22*(5), bbab009. doi:10.1093/bib/bbab009