Explainable Artificial Intelligence for Predicting Oncogenic Gene Expression Programs Driven by MYC Condensate Dynamics

Shane Adams
Department of Computer Science, George Mason University, Fairfax, VA, USA.
shane1999@gmu.edu

Abstract

The emergence of phase separation as a fundamental organizing principle in transcriptional regulation has reshaped our understanding of oncogenic gene expression, particularly for the MYC oncoprotein whose condensate dynamics orchestrate selective transcriptional programs. Concurrently, explainable artificial intelligence has become indispensable for interpreting high-dimensional genomic data and for building trust in predictive models used in clinical and research settings. This paper presents a systems-level analysis of an explainable AI framework designed to predict oncogenic gene expression programs driven by MYC condensate dynamics. We examine the architectural trade-offs between predictive accuracy and interpretability, the robustness of model explanations under biological variability, and the infrastructural requirements for deploying such systems in translational workflows. Governance and fairness considerations are discussed through the lens of algorithmic bias in genomics, reproducibility of explanations, and ethical implications for personalized oncology. Cross-domain comparisons with other biophysical systems, such as transcriptional condensates at super-enhancers, highlight the generality of the framework. Sustainability of model training, data provenance, and the policy landscape for responsible AI in molecular biology are also addressed. The paper argues that while explainable AI offers powerful tools for deciphering complex phase-separation-driven transcription, its effective integration demands careful attention to system architecture, validation protocols, and socio-technical governance to ensure reliable and equitable outcomes.

Keywords

explainable artificial intelligence, MYC condensate dynamics, oncogenic gene expression, phase separation, system architecture, algorithmic governance, robustness, fairness, infrastructure.

1. Introduction

The discovery that many transcription factors and coactivators form dynamic, liquid-like condensates through phase separation has fundamentally altered the conceptual framework for understanding transcriptional control in eukaryotic cells [6,7,8]. Among the most clinically relevant examples is the MYC oncoprotein, whose aberrant expression drives a wide spectrum of human cancers. MYC functions as a global amplifier of transcription, yet its activity is not indiscriminate; it selectively modulates gene expression programs that promote cell proliferation, metabolism, and genomic instability [9,10]. Recent experimental evidence has demonstrated that MYC phase separation selectively modulates the transcriptome, establishing a direct mechanistic link between condensate formation and the activation of specific oncogenic gene expression programs [12]. This insight opens new avenues for

therapeutic intervention but also poses significant computational challenges: predicting which sets of genes will be influenced by MYC condensate dynamics under varying cellular contexts requires models that can integrate biophysical principles with high-throughput genomic data.

Machine learning, particularly deep learning, has achieved remarkable success in predicting gene expression from sequence and epigenetic features [19,20]. However, these models often operate as black boxes, making it difficult to understand the biological rationale behind their predictions. In high-stakes applications such as oncology, where treatment decisions may be guided by predicted gene expression programs, interpretability is not merely a convenience but a necessity [1,4]. Explainable artificial intelligence (XAI) has emerged as a critical subfield that aims to render model outputs transparent, providing human-understandable explanations for individual predictions [5]. For systems that predict the consequences of MYC condensate dynamics, XAI methods can reveal which sequence features, chromatin states, or biophysical parameters most strongly influence the predicted gene expression outcomes, thereby enabling biological validation and hypothesis generation.

This paper adopts a systems-level perspective to examine the design, deployment, and governance of XAI frameworks for predicting oncogenic gene expression programs driven by MYC condensate dynamics. Rather than focusing on a specific algorithmic implementation, we analyze the structural trade-offs inherent in such systems, including the tension between model complexity and interpretability, the robustness of explanations to biological and technical noise, and the infrastructural requirements for large-scale deployment in clinical and research environments. We also consider the broader socio-technical dimensions: fairness in model performance across diverse genetic backgrounds, the reproducibility of explanations, and the policy implications of using AI-driven predictions in precision oncology. By drawing on cross-domain comparisons with similar challenges in other biophysical systems and by referencing established principles from algorithmic governance, we aim to provide a comprehensive framework for building trustworthy and sustainable predictive systems in the era of phase-separation biology.

2. Biological Mechanisms of MYC Condensate Dynamics and Transcriptional Regulation

Phase separation is now recognized as a pervasive mechanism for compartmentalizing biochemical reactions within cells without the need for membrane boundaries [11]. In the nucleus, transcription factors and coactivators with intrinsically disordered regions can undergo liquid-liquid phase separation to form condensates that concentrate RNA polymerase II and other transcriptional machinery at specific genomic loci, such as super-enhancers [7,8]. MYC, a basic helix-loop-helix transcription factor, contains an N-terminal transactivation domain that is rich in intrinsically disordered sequences, enabling it to participate in phase separation [12]. The formation of MYC condensates is highly sensitive to its concentration, post-translational modifications, and interactions with partner proteins, all of which can be altered in cancer.

The functional consequence of MYC condensate formation is the selective activation of a subset of target genes that are enriched for promoters and enhancers with specific sequence and chromatin signatures. Unlike earlier models that posited MYC as a general amplifier of all active transcription, the phase-separation model suggests that MYC condensates preferentially engage with loci that have a high density of cognate E-box motifs and that are already marked by active histone modifications [10,12]. This selectivity implies that

predicting the transcriptional output of MYC condensate dynamics requires not only knowledge of MYC binding sites but also an integrated assessment of chromatin state, three-dimensional genome architecture, and the local biophysical environment that influences condensate stability.

High-throughput experimental methods, such as chromatin immunoprecipitation sequencing, RNA sequencing, and proximity ligation assays, generate vast amounts of data that can be used to train predictive models. However, these data are inherently noisy, sparse in coverage of rare cell states, and often collected from bulk populations that mask cell-to-cell variability. Moreover, the dynamic nature of condensates—their formation, dissolution, and material properties—introduces a temporal dimension that static genomic assays cannot capture. Therefore, any predictive system must be designed to handle missing data, incorporate mechanistic priors, and provide explanations that are robust to measurement uncertainty. The complexity of these interactions underscores the need for XAI methods that can attribute predictions to specific biological features in a manner that aligns with experimental validation.

3. Explainable AI Frameworks for Gene Expression Prediction

Addressing the challenge of predicting gene expression from MYC condensate dynamics requires a modeling pipeline that integrates feature engineering, representation learning, and interpretability. Deep learning architectures, such as convolutional neural networks and transformer models, have been successfully applied to sequence-based prediction of transcription factor binding and chromatin accessibility [19,20]. These models learn hierarchical patterns from raw DNA sequence and can incorporate additional modalities such as chromatin immunoprecipitation signals, histone modification marks, and three-dimensional contact maps. For the specific case of MYC condensate-driven transcription, the model must also incorporate features related to phase separation propensity, such as the density of intrinsically disordered regions, the predicted saturation concentration of MYC, and the presence of scaffold proteins that promote condensate formation.

XAI techniques can be broadly categorized into intrinsically interpretable models and post-hoc explanation methods. Intrinsically interpretable models, such as linear models or decision trees, offer transparency by design but often lack the capacity to capture complex nonlinear interactions [1]. In contrast, post-hoc methods like SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) can be applied to any black-box model to produce feature attribution scores for individual predictions [2,3]. For genomic applications, SHAP has been particularly useful because it provides a unified measure of feature importance that satisfies desirable properties such as consistency and local accuracy. When applied to a deep learning model trained to predict the expression of MYC target genes, SHAP can identify which sequence motifs, chromatin marks, or condensate-related features most strongly contribute to a given prediction, thereby providing a mechanistic hypothesis for experimental follow-up.

However, the use of post-hoc explanations introduces its own set of challenges. Explanations can be unstable across different input perturbations, leading to false confidence in the model's reasoning [4]. In the context of MYC condensate dynamics, biological variability—such as differences in MYC concentration across cell types or the presence of competing phase-separating factors—can cause the same model to produce different explanations for similar inputs. Methods that measure the robustness of explanations, such as adversarial perturbation testing or sensitivity analysis, are therefore essential for assessing the reliability of the system. Furthermore, the choice of baseline values in SHAP can dramatically affect the attribution

results, and domain-specific knowledge is required to select biologically meaningful baselines. These considerations highlight the need for rigorous validation protocols that go beyond standard performance metrics and include explanation fidelity tests.

4. System-Level Design and Architectural Considerations

Deploying an XAI system for predicting MYC condensate-driven transcription at scale requires careful engineering of the data infrastructure, model training pipeline, and explanation delivery mechanism. The data pipeline must handle heterogeneous data types—genomic sequences, epigenomic profiles, proteomic measurements, and biophysical parameters of condensates—that are often stored in different formats and accessed through distinct databases. A unified data warehouse with standardized ontologies and metadata is essential for ensuring reproducibility and enabling cross-study comparisons. The volume of raw sequencing data is enormous, often reaching hundreds of gigabytes per experiment, and preprocessing steps such as alignment, peak calling, and normalization must be automated and version-controlled to prevent errors from propagating into the model.

Model training presents both computational and methodological challenges. The number of available training samples, defined as cell types or conditions for which both MYC condensate measurements and transcriptomic outcomes are available, is limited. Transfer learning, where a model pre-trained on large-scale genomic datasets is fine-tuned on the smaller MYC-specific dataset, can alleviate data scarcity. However, transfer learning introduces its own risks of negative transfer if the source domain is not sufficiently related. Architectural choices, such as the depth of the network or the inclusion of attention mechanisms, must balance the need for expressive power with the risk of overfitting. Regularization techniques, including dropout and early stopping, are standard, but additional domain-specific constraints—such as enforcing that the model respects known regulatory hierarchies—can improve generalization.

The explanation delivery system must be designed with the end user in mind. For a computational biologist, a global feature importance plot may suffice to identify the most relevant genomic features across all genes. For a clinician considering a therapeutic recommendation, a local explanation for a specific patient's tumor sample is required, and this explanation must be presented in an intuitive manner without oversimplifying the underlying biology. Building an interactive dashboard that allows users to probe predictions, explore alternative scenarios, and compare explanations across models is a promising approach. The system must also log all predictions and explanations for audit purposes, enabling retrospective analysis of model behavior and supporting regulatory compliance in clinical settings.

5. Trade-Offs Between Accuracy, Interpretability, and Robustness

A central tension in any XAI system is the trade-off between predictive accuracy and interpretability. In the domain of gene expression prediction, deep neural networks often achieve higher accuracy than simpler models because they can capture complex nonlinear dependencies and long-range interactions in genomic sequences [19]. However, these models are inherently opaque, and post-hoc explanations may not faithfully represent the model's internal logic. Research has shown that explanations can be misleading when the model relies on spurious correlations or when the explanation method itself is biased [1]. This is particularly problematic in oncology, where a treatment decision based on a flawed explanation could have severe consequences.

One approach to mitigating this trade-off is to use concept-based explanations, where the model is trained to predict biologically meaningful intermediate representations (such as transcription factor binding events or chromatin state annotations) before making the final gene expression prediction. These intermediate concepts are inherently interpretable, and the model's reasoning can be traced through them. For MYC condensate dynamics, concepts could include the intensity of MYC binding at a promoter, the local degree of chromatin condensation, or the estimated saturation concentration of MYC condensates. By enforcing that the model's predictions arise from these concepts, the system gains interpretability without necessarily sacrificing accuracy, as the concepts themselves may be predictive.

Robustness is another critical dimension. A model that achieves high accuracy on a held-out test set may fail when applied to a new tissue type or a patient with a rare genetic variant that alters MYC condensate properties. Domain adaptation techniques, such as adversarial training or invariant risk minimization, can improve out-of-distribution generalization, but they often require additional data or assumptions about the sources of covariate shift. Explanations must also be robust: a small perturbation in the input should not drastically change the feature attribution. Quantifying explanation robustness through metrics like the maximum sensitivity to input perturbations allows researchers to flag predictions that are not trustworthy. In the clinical pipeline, such predictions could be sent for manual review rather than automatically acted upon.

6. Governance, Fairness, and Ethical Implications

The use of AI to predict oncogenic gene expression programs raises significant governance and fairness concerns that extend beyond technical performance. One key issue is algorithmic bias: if the training data are predominantly derived from cell lines or patient populations of European ancestry, the model may perform poorly on individuals from underrepresented groups, leading to disparities in the accuracy and reliability of predictions [16,17]. In the context of MYC condensate dynamics, genetic variation in regulatory regions, as well as differences in the expression levels of MYC or its binding partners across populations, could influence condensate formation and, consequently, the gene expression programs that the model aims to predict. Without careful validation across diverse ancestries, the system may exacerbate existing health inequities.

Fairness in XAI also concerns the explanations themselves. If the model provides explanations that are systematically less accurate or less informative for certain subgroups, then users from those subgroups may be disadvantaged in their ability to understand and trust the predictions. For example, if the model's explanation for a tumor sample from a patient with a rare MYC mutation relies on features that are not well represented in the training data, the explanation may be misleading. Algorithmic auditing, where independent evaluators assess model performance and explanation quality across predefined subgroups, is a necessary component of responsible deployment.

Policy implications are profound. As AI systems become integrated into clinical decision support for precision oncology, regulatory bodies such as the U.S. Food and Drug Administration and the European Medicines Agency will need to establish guidelines for validating and monitoring these systems. The dynamic nature of phase-separation biology, combined with the evolving understanding of condensate behavior, means that models may need to be updated frequently. Governance frameworks must therefore address model lifecycle management, including versioning, retraining triggers, and the process for withdrawing a model from use if its performance degrades. Transparency obligations should

require that the training data, model architecture, and explanation methods are publicly documented to enable scientific scrutiny [18].

7. Deployment Infrastructure and Sustainability

Deploying an XAI system for MYC condensate-driven gene expression prediction in a real-world setting demands robust computational infrastructure. High-performance computing clusters or cloud-based platforms are necessary to handle the training of deep neural networks on genomic-scale data, as well as the real-time inference and explanation generation for individual queries. Containerization technologies, such as Docker and Kubernetes, facilitate reproducible deployments across different environments, while workflow management systems like Nextflow or Snakemake orchestrate complex data processing pipelines. The system must also interface with existing laboratory information management systems and electronic health records, requiring adherence to data privacy regulations such as HIPAA and GDPR.

Sustainability is an often-overlooked aspect of large-scale AI systems. Training deep learning models on genomic data consumes significant energy, contributing to the carbon footprint of research institutions. Model compression techniques, such as pruning, quantization, and knowledge distillation, can reduce the computational cost of inference without substantial loss of accuracy. Additionally, using pre-trained models and fine-tuning on smaller datasets, as mentioned earlier, reduces the need for repeated full training runs. Sustainable software engineering practices, such as efficient data loading and optimized tensor operations, should be prioritized. Beyond environmental sustainability, the system must be financially sustainable: maintenance, updates, and personnel costs must be accounted for in long-term funding strategies, particularly for academic or public health deployments.

8. Conclusion

The convergence of phase-separation biology and explainable artificial intelligence offers a powerful approach to understanding and predicting the oncogenic gene expression programs driven by MYC condensate dynamics. This paper has examined the system-level challenges associated with building such predictive frameworks, emphasizing the architectural trade-offs between accuracy, interpretability, and robustness, as well as the infrastructural, governance, and ethical dimensions that are essential for responsible deployment. While deep learning models and post-hoc explanation methods provide the technical core of the system, their effectiveness is contingent on careful design choices that account for biological variability, data limitations, and the needs of diverse end users. Cross-domain comparisons with similar efforts in other biophysical systems highlight both the promise and the pitfalls of applying XAI to complex molecular processes. As the field moves toward clinical translation, sustained attention to fairness, sustainability, and regulatory alignment will be critical to ensuring that these technologies benefit all patients equitably. Future research should focus on developing explanation methods that are inherently robust to distributional shift, integrating mechanistic models of condensate physics with data-driven approaches, and establishing community standards for benchmarking and validation. Only through such interdisciplinary collaboration can we unlock the full potential of explainable AI in the era of phase-separation-driven transcription.

References

1. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206-215.

2. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30.

3. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135-1144.

4. Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.

5. Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82-115.

6. Boija, A., Klein, I. A., Sabari, B. R., Dall'Agnese, A., Coffey, E. L., Zamudio, A. V., ... & Young, R. A. (2018). Transcription factors activate genes through the phase-separation capacity of their activation domains. Cell, 175(7), 1842-1855.

7. Sabari, B. R., Dall'Agnese, A., Boija, A., Klein, I. A., Coffey, E. L., Shrinivas, K., ... & Young, R. A. (2018). Coactivator condensation at super-enhancers links phase separation and gene control. Science, 361(6400), eaar3958.

8. Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K., & Sharp, P. A. (2017). A phase separation model for transcriptional control. Cell, 169(1), 13-23.

9. Kress, T. R., Sabò, A., & Amati, B. (2015). MYC: connecting selective transcriptional control to global RNA production. Nature Reviews Cancer, 15(10), 593-607.

10. Lin, C. Y., Lovén, J., Rahl, P. B., Paranal, R. M., Burge, C. B., Bradner, J. E., ... & Young, R. A. (2012). Transcriptional amplification in tumor cells with elevated c-Myc. Cell, 151(1), 56-67.

11. Shin, Y., & Brangwynne, C. P. (2017). Liquid phase condensation in cell physiology and disease. Science, 357(6357), eaaf4382.

12. Yang, J., Chung, C. I., Koach, J., Liu, H., Navalkar, A., He, H., ... & Shu, X. (2024). MYC phase separation selectively modulates the transcriptome. Nature Structural & Molecular Biology, 31(10), 1567-1579.

13. Chen, D., & Liu, Y. (2023). Deep learning for gene expression prediction: a review. Briefings in Bioinformatics, 24(1), bbac536.

14. Ching, T., Himmelstein, D. S., Beaulieu-Jones, B. K., Kalinin, A. A., Do, B. T., Way, G. P., ... & Greene, C. S. (2018). Opportunities and obstacles for deep learning in biology and medicine. Journal of the Royal Society Interface, 15(141), 20170387.

15. Yu, M. K., Ma, J., & Fisher, J. (2020). Using mechanistic models to integrate - not replace - machine learning and mechanistic biology. Nature Machine Intelligence, 2(6), 307-309.

16. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Vayena, E. (2018). AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. Minds and Machines, 28, 689-707.

17. Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2), 2053951716679679.

18. Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389-399.

19. Angermueller, C., Pärnamaa, T., Parts, L., & Stegle, O. (2016). Deep learning for computational biology. Molecular Systems Biology, 12(7), 878.

20. Eraslan, G., Avsec, Ž., Gagneur, J., & Theis, F. J. (2019). Deep learning: new computational modelling techniques for genomics. Nature Reviews Genetics, 20(7), 389-403.