Concept Paper



# Artificial Intelligence as a Conceptual Lens on Gene Latency and Genomic Memory

Abdulmohsen H. Alrohaimi *

Posted Date: 5 March 2026

doi: 10.20944/preprints202603.0454.v1

Keywords: gene latency; genomic memory; artificial intelligence; transformer models; latent representations; systems biology; computational genomics

 doi:10.20944/preprints202603.0454.v1



*Concept Paper*

# Artificial Intelligence as a Conceptual Lens on Gene Latency and Genomic Memory

**Abdulmohsen H. Alrohaimi**

Department of Pharmacology, College of Pharmacy, Shaqra University, Saudi Arabia; alrohaimi@su.edu.sa

**Abstract**

Genomic regulation is typically interpreted through observable molecular states such as gene expression, chromatin accessibility and epigenetic modifications. However, biological systems also contain large reservoirs of genomic information that remain transcriptionally inactive for extended periods while retaining the capacity to influence future regulatory behaviour. This phenomenon, referred to here as **gene latency**, suggests that genomes may preserve forms of biological memory beyond currently expressed molecular states. Recent advances in artificial intelligence—particularly transformer-based architectures—demonstrate how complex systems can encode structured information within latent representational spaces that influence outputs without continuous activation (Vaswani et al., 2017; Brown et al., 2020). In this study, we propose a conceptual framework that interprets gene latency as a form of **genomic memory** using principles derived from latent representation learning in artificial intelligence. By aligning concepts from systems biology, epigenetics and machine learning, we outline theoretical and computational perspectives for identifying latent regulatory potential within genomic systems. This framework suggests that genomes may contain distributed reservoirs of regulatory capacity shaped by developmental history and environmental exposure. Integrating artificial intelligence with genomic theory may therefore enable new approaches for modelling latent regulatory states and predicting transitions from genomic latency to activation.

**Keywords:** gene latency; genomic memory; artificial intelligence; transformer models; latent representations; systems biology; computational genomics

## 1. Introduction

The regulation of gene activity is central to biological function and cellular identity. Advances in genomic technologies—including RNA sequencing, chromatin immunoprecipitation sequencing and single-cell multi-omics—have enabled high-resolution measurement of gene expression and regulatory activity across diverse biological contexts (ENCODE Project Consortium, 2012; Roadmap Epigenomics Consortium, 2015). These approaches have significantly expanded understanding of transcriptional regulation and gene regulatory networks.

Despite these advances, many models of genomic regulation remain fundamentally **state-based**, focusing primarily on observable molecular outputs such as transcription levels or epigenetic modifications (Cavalli and Heard, 2019). However, biological systems frequently display behaviours that cannot be fully explained by current molecular states alone. Processes such as development, aging and chronic disease progression often reflect long-term influences of past regulatory states on present cellular behaviour (López-Otín et al., 2013; López-Otín et al., 2023).

These observations raise a central question in systems biology: **how do genomes preserve regulatory potential across time?**

One possible explanation lies in the concept of **gene latency**, defined here as genomic information that remains transcriptionally inactive yet retains the capacity to influence future regulatory behaviour. Latent genomic elements may include silenced genes, pseudogenes and regulatory sequences whose functional potential persists despite the absence of current expression.

 doi:10.20944/preprints202603.0454.v1



Previous work has explored related ideas through the study of epigenetic memory, chromatin states and dormant regulatory elements (Bird, 2007; Bonasio, Tu and Reinberg, 2010; Ernst and Kellis, 2010). However, these phenomena have rarely been integrated into a unified conceptual framework describing how genomes preserve regulatory potential over time. To our knowledge, this work represents one of the first attempts to frame genomic inactivity as a structured form of biological memory through the concept of **gene latency**.

In parallel, artificial intelligence research has produced systems capable of storing large amounts of contextual knowledge within **latent representational spaces**. Transformer-based neural networks encode complex patterns within distributed embeddings that remain inactive until activated by specific contextual prompts (Vaswani et al., 2017; Devlin et al., 2019; Brown et al., 2020).

These developments raise the possibility that artificial intelligence may provide a useful conceptual framework for understanding how biological systems store and retrieve latent regulatory information.

## 2. Materials and Methods

*Conceptual Modelling Framework*

This study develops a theoretical framework integrating concepts from genomics, epigenetics and machine learning in order to examine gene latency as a form of genomic memory.

The framework is constructed through comparative analysis of:

1. Biological regulatory systems
2. Latent representation learning in artificial intelligence

The objective is not to propose mechanistic equivalence between biological and artificial systems but to explore conceptual parallels that may guide future computational approaches.

*Literature Synthesis*

Relevant literature from the following domains was systematically reviewed:

- epigenetics and chromatin biology
- systems genomics
- machine learning and deep learning
- computational genomics

Foundational studies on genomic regulation (Felsenfeld and Groudine, 2003; Levine and Tjian, 2003), epigenetic inheritance (Ptashne, 2013; Cavalli and Heard, 2019) and aging biology (López-Otín et al., 2013; López-Otín et al., 2023) were analysed alongside major advances in artificial intelligence and deep learning (LeCun, Bengio and Hinton, 2015; Goodfellow, Bengio and Courville, 2016).

*Conceptual Mapping Approach*

Analogies between biological regulatory systems and artificial intelligence architectures were mapped across three conceptual dimensions:

1. information storage
2. contextual activation
3. distributed representation

These dimensions were used to construct a conceptual model linking genomic latency with latent representation learning.

## 3. Results

*Gene Latency as a Reservoir of Genomic Information*

Biological genomes contain numerous sequences that remain transcriptionally inactive across extended periods yet remain structurally preserved. These include pseudogenes, dormant regulatory

 doi:10.20944/preprints202603.0454.v1



elements and silenced gene clusters (Pink et al., 2011). Although these elements may not produce immediate molecular outputs, they may retain regulatory potential that becomes relevant under specific environmental or developmental conditions.

*Parallels with Latent Representation Learning*

Transformer-based neural networks encode knowledge within high-dimensional latent embeddings learned during training (Vaswani et al., 2017). These embeddings capture contextual relationships that influence system behaviour without requiring continuous activation.

Similarly, latent genomic elements may represent biological embeddings shaped by developmental history and epigenetic constraints.

*Distributed Genomic Memory*

Genomic regulation often emerges from distributed interactions across multiple regulatory layers, including enhancers, chromatin loops and transcription factor networks (Dekker and Mirny, 2016). This distributed architecture parallels attention-based mechanisms in transformer models, where outputs depend on weighted interactions across multiple representational layers.

Recent developments in deep learning have demonstrated the capacity of machine learning models to capture complex biological patterns within genomic sequences (Alipanahi et al., 2015; Zhou and Troyanskaya, 2015; Koo and Eddy, 2019; Avsec et al., 2021). The rapid expansion of artificial intelligence technologies—including protein structure prediction systems such as AlphaFold—further illustrates the growing role of machine learning in biological discovery (Senior et al., 2020; Jumper et al., 2021; Silver et al., 2016).

## 4. Discussion

The concept of gene latency suggests that genomes may store regulatory potential beyond what is currently observable through molecular assays. Traditional genomic analyses focus primarily on active regulatory states, yet biological systems frequently exhibit behaviours that imply the presence of long-term regulatory memory.

Artificial intelligence provides a useful conceptual analogy because modern machine learning models demonstrate how complex systems can store structured information in latent form within internal representational spaces (LeCun, Bengio and Hinton, 2015; Goodfellow, Bengio and Courville, 2016).

This perspective does not imply that genomes function like neural networks. Rather, it highlights a conceptual framework through which biological regulatory systems may be interpreted as preserving information across time in ways analogous to latent representation learning.

Future research may explore computational approaches capable of identifying latent regulatory capacity within genomic systems, particularly through longitudinal single-cell datasets and deep learning models trained on genomic sequences.

## 5. Conclusion

Genomes may contain extensive reservoirs of latent regulatory information that remain invisible to traditional state-based analyses. Interpreting gene latency through the conceptual framework of latent representation learning provides a new perspective on genomic organisation.

By integrating insights from artificial intelligence and systems biology, researchers may develop new computational methods for identifying genomic memory and predicting transitions from latent to active regulatory states. Future empirical and computational studies will be required to determine whether gene latency can be quantitatively characterised as a measurable dimension of genomic organisation.

 doi:10.20944/preprints202603.0454.v1



**Author Contributions:** Abdulmohsen H. Alrohaimi conceived the conceptual framework, conducted the literature synthesis and wrote the manuscript.

**Acknowledgments:** The author thanks colleagues in computational biology and artificial intelligence for valuable discussions.

**Competing Interests:** The author declares no competing interests.

## References

1. Alipanahi, B. et al., 2015. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nature Biotechnology*, 33, pp.831–838.
2. Avsec, Ž. et al., 2021. Effective gene expression prediction from sequence by integrating long-range interactions. *Nature Methods*, 18, pp.1196–1203.
3. Bird, A., 2007. Perceptions of epigenetics. *Nature*, 447, pp.396–398.
4. Bonasio, R., Tu, S. and Reinberg, D., 2010. Molecular signals of epigenetic states. *Science*, 330, pp.612–616.
5. Brown, T. et al., 2020. Language models are few-shot learners. *NeurIPS*, 33, pp.1877–1901.
6. Cavalli, G. and Heard, E., 2019. Advances in epigenetics link genetics to the environment and disease. *Nature*, 571, pp.489–499.
7. Chen, K. and Rajewsky, N., 2007. The evolution of gene regulation. *Nature Reviews Genetics*, 8, pp.93–103.
8. Dekker, J. and Mirny, L., 2016. The 3D genome as moderator of chromosomal communication. *Cell*, 164, pp.1110–1121.
9. Devlin, J. et al., 2019. BERT: Pre-training of deep bidirectional transformers. *NAACL-HLT*, pp.4171–4186.
10. ENCODE Project Consortium, 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489, pp.57–74.
11. Eraslan, G. et al., 2019. Deep learning for genomics. *Nature Reviews Genetics*, 20, pp.389–403.
12. Ernst, J. and Kellis, M., 2010. Chromatin-state discovery. *Nature Biotechnology*, 28, pp.817–825.
13. Felsenfeld, G. and Groudine, M., 2003. Controlling the double helix. *Nature*, 421, pp.448–453.
14. Goodfellow, I., Bengio, Y. and Courville, A., 2016. *Deep Learning*. MIT Press.
15. Jumper, J. et al., 2021. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596, pp.583–589.
16. Kingma, D. and Welling, M., 2014. Auto-encoding variational Bayes. *ICLR*.
17. Koo, P. and Eddy, S., 2019. Representation learning of genomic sequence motifs. *Nature Reviews Genetics*, 20, pp.611–624.
18. LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. *Nature*, 521, pp.436–444.
19. Levine, M. and Tjian, R., 2003. Transcription regulation and animal diversity. *Nature*, 424, pp.147–151.
20. López-Otín, C. et al., 2013. The hallmarks of aging. *Cell*, 153, pp.1194–1217.
21. López-Otín, C. et al., 2023. Hallmarks of aging: an expanding universe. *Cell*, 186, pp.243–278.
22. Pink, R. et al., 2011. Pseudogenes: regulators in health and disease. *RNA*, 17, pp.792–798.
23. Roadmap Epigenomics Consortium, 2015. Integrative analysis of reference human epigenomes. *Nature*, 518, pp.317–330.
24. Senior, A. et al., 2020. Improved protein structure prediction using deep learning. *Nature*, 577, pp.706–710.
25. Silver, D. et al., 2016. Mastering the game of Go with deep neural networks. *Nature*, 529, pp.484–489.
26. Vaswani, A. et al., 2017. Attention is all you need. *NeurIPS*, 30, pp.5998–6008.
27. Zhou, J. and Troyanskaya, O., 2015. Predicting effects of noncoding variants. *Nature Methods*, 12, pp.931–934.