Journal of Carcinogenesis
Covering Genesis to Therapeutics

Journal of Carcinogenesis, ISSN: 1477-3163
2024; Vol 23: Issue 1, https://carcinogenesis.com/

Genome-Wide Computational Prediction And Analysis Of Putative Promoters And Transcription Site: Derived In All Locus Of Whole Genome Present In Mus Musculus

Samreen[1], Dinesh Prajapati[2], Shashank Upadhyay[3], Mohd. Faheem Khan[4], Pallavi Sharma[5]

[1,2,3]Department of Biotechnology, Invertis University, Bareilly, U.P, India.

[4]Department of Biotechnology, Khandelwal College of Management Science and Technology, Bareilly, U.P, India.

[5]R&D Division, Somics Lifesciences Pvt. Ltd. Bareilly, U.P, India.

ABSTRACT

Dysregulation of transcriptional control is a central molecular event in carcinogenesis and neurodegenerative disorders, frequently arising from aberrant promoter architecture and altered transcription start site (TSS) usage. Genome-wide identification of promoter regions therefore provides critical insight into gene regulatory mechanisms underlying disease susceptibility. In the present study, we performed a comprehensive in silico analysis of the complete Mus musculus genome (GRCm39) to systematically predict and characterize putative RNA polymerase II promoters and transcription start sites across all genomic loci.

Whole-genome sequences were retrieved from the NCBI repository and analyzed using an integrated bioinformatics framework incorporating GeneRunner, DNASTAR, ExPASy, and the BDGP neural-network-based Promoter 2.0 algorithm. Chromosome-wise variation in genome size, GC content, and gene composition was evaluated, followed by strand-specific promoter prediction and transcription factor site enrichment analysis. Our results reveal extensive heterogeneity in promoter distribution across chromosomes, with significant enrichment of GC-rich and TATA-like motifs within upstream regulatory regions. Several high-confidence transcription start site hotspots were identified, indicating preferential loci of transcriptional initiation.

Notably, the predicted promoter regions exhibit architectural features consistent with estrogen receptor–responsive elements, suggesting potential regulatory relevance to hormone-dependent gene expression and carcinogenic pathways. Advanced graphical representations, including promoter density mapping and GC-content correlation analyses, further elucidated the structural determinants of transcriptional regulation. Collectively, this genome-wide promoter atlas provides a valuable computational resource for functional genomics and establishes a regulatory framework for investigating estrogen-mediated transcriptional dysregulation in cancer and neurological disorders.

Keywords: *Mus musculus, promoter prediction, transcription start site, estrogen response element, gene regulation, carcinogenesis*

How to Cite: Samreen, Dinesh Prajapati, Shashank Upadhyay, Mohd. Faheem Khan, Pallavi Sharma, (2024) Genome-Wide Computational Prediction And Analysis Of Putative Promoters And Transcription Site: Derived In All Locus Of Whole Genome Present In Mus Musculus, *Journal of Carcinogenesis, Vol.23, No.1*, 562-570

1. INTRODUCTION

Precise regulation of gene expression is fundamental to cellular homeostasis, development, and tissue specificity, whereas its dysregulation constitutes a hallmark of carcinogenesis and neurological disorders (1,2). Transcriptional control is primarily governed by promoter regions and transcription start sites (TSSs), which integrate signals from transcription factors, epigenetic modifications, and chromatin architecture to initiate RNA polymerase II–mediated transcription (3). Aberrant promoter usage altered TSS selection, and epigenetic remodeling of regulatory regions have been widely implicated in oncogene activation and tumor suppressor gene silencing (4).

Steroid hormones, particularly estrogens, exert profound effects on transcriptional regulation through estrogen receptors (ERs), which function as ligand-activated transcription factors (5). Estrogens influence diverse biological processes ranging

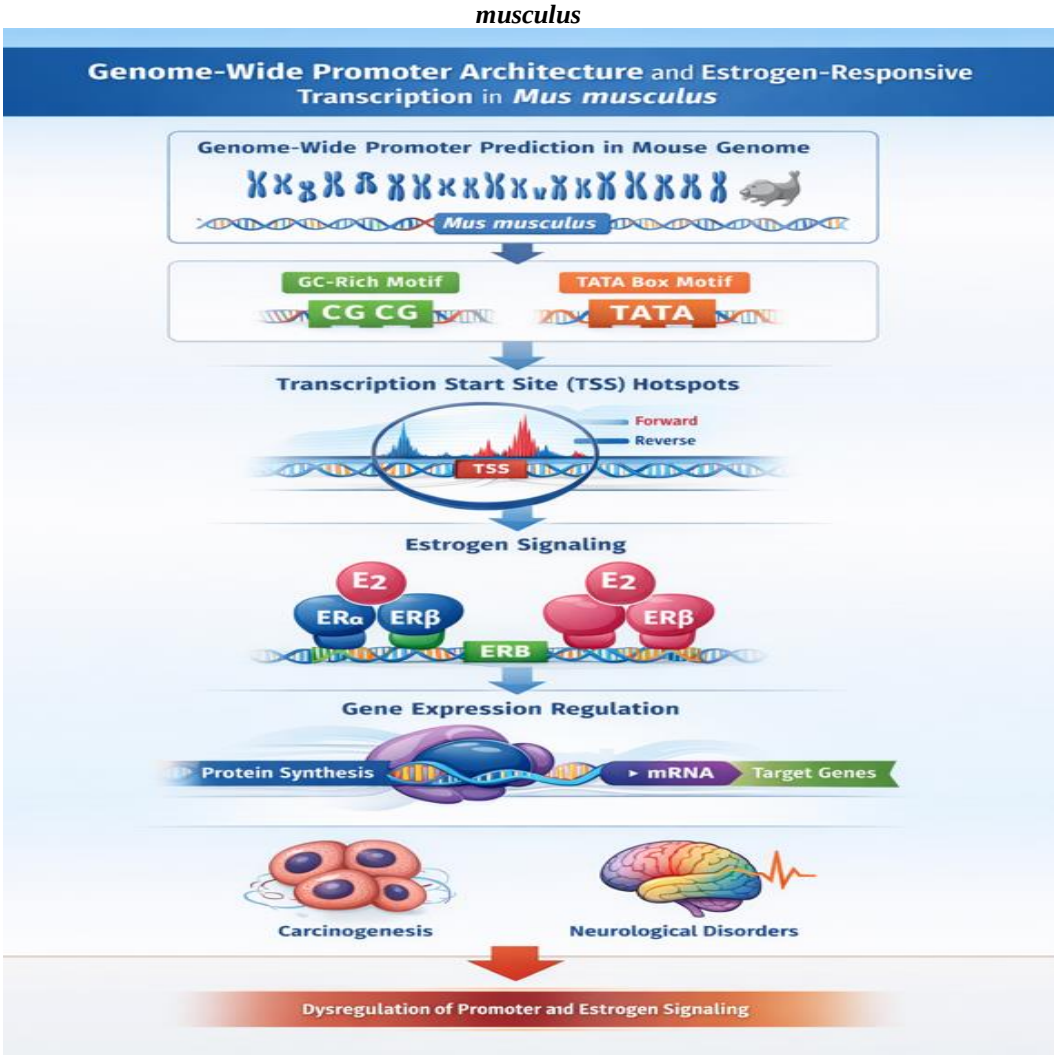
from reproductive physiology to cognition, neuronal differentiation, and cell proliferation (6,7). Accumulating evidence demonstrates that estrogen signaling plays a pivotal role in hormone-dependent cancers, including breast, ovarian, and endometrial malignancies, as well as in neurodegenerative conditions such as Alzheimer's disease (8,9).

Estrogen receptors exist predominantly as two isoforms, ERα and ERβ, encoded by the *ESR1* and *ESR2* genes, respectively (10). These receptors regulate gene expression via both genomic and non-genomic mechanisms. In the classical genomic pathway, estrogen-bound ERs directly interact with estrogen response elements (EREs), consensus DNA motifs located within promoter or enhancer regions of target genes (11). Even single nucleotide variations within ERE consensus sequences can significantly alter ER binding affinity, thereby modulating transcriptional output and disease susceptibility (12).

Beyond reproductive tissues, estrogen signaling has emerged as a critical regulator of brain development and function. ERα and ERβ exhibit region- and cell-type-specific expression patterns within the central nervous system, influencing synaptic plasticity, memory formation, stress response, and neuroprotection (13,14). Importantly, dysregulated estrogen signaling has been linked to sex-specific vulnerability to neurodegenerative diseases and gliomas, further underscoring its relevance to carcinogenesis and neurological pathology (15).

The laboratory mouse (*Mus musculus*) serves as a premier mammalian model for investigating transcriptional regulation due to its high genomic homology with humans, well-annotated genome assemblies, and extensive availability of genetic and molecular tools (16). Genome-wide computational analyses in *Mus musculus* have proven invaluable for identifying regulatory elements, elucidating transcriptional networks, and modeling disease-associated gene expression patterns (17). A schematic overview of the genome-wide promoter prediction strategy and estrogen receptor–mediated transcriptional regulation is presented in the graphical abstract as in figure 1.

Figure 1: Genome-Wide Promoter Architecture and Estrogen-Responsive Transcriptional Regulation in *Mus musculus*



Despite significant advances in transcriptomics and epigenomics, a comprehensive genome-wide map of putative promoters and transcription start sites in *Mus musculus* remains essential for integrative studies of hormone-responsive and cancer-associated transcriptional regulation. In particular, systematic annotation of promoter architecture provides a foundational framework for understanding how estrogen-mediated signaling interfaces with transcriptional dysregulation in carcinogenesis. Therefore, the present study aims to perform a genome-wide *in silico* prediction and analysis of putative promoters and transcription start sites across all loci of the *Mus musculus* genome, thereby generating a regulatory atlas with direct relevance to cancer biology and neurogenomic research (18).

2. MATERIALS AND METHODS

2.1 Genome Data Retrieval

The complete reference genome sequence of *Mus musculus* (assembly GRCm39) was retrieved from the National Center for Biotechnology Information (NCBI) genome repository (https://www.ncbi.nlm.nih.gov/genome). Chromosome-wise FASTA files, along with corresponding annotation metadata, were downloaded to ensure comprehensive coverage of autosomes, sex chromosomes, mitochondrial DNA, and unplaced genomic scaffolds. Only curated reference sequences were included to maintain data integrity and analytical consistency.

2.2 Genome Composition and GC Content Analysis

Genome size, nucleotide distribution, and GC content were analyzed on a chromosome-wise basis using an integrated bioinformatics workflow. GeneRunner (Version 6.5), DNASTAR Lasergene suite, and ExPASy bioinformatics tools were employed to calculate genome length, GC percentage, and base composition. GC content was computed using sliding window analysis to assess regional variability and its association with promoter enrichment. These parameters were summarized statistically and used for downstream correlation analysis with predicted promoter density.

2.3 Promoter Prediction Using Neural Network Algorithm

Putative RNA polymerase II promoter regions were identified using Promoter 2.0, a BDGP neural-network-based promoter prediction tool specifically designed for vertebrate genomes. The algorithm integrates transcription factor binding site density, promoter-specific sequence features, TATA-box scoring matrices, and non-promoter background models. Promoter prediction was performed independently for forward and reverse strands to capture strand-specific regulatory elements.

Promoter regions were defined within upstream and downstream flanking sequences of annotated genes, with prediction scores above the recommended threshold considered significant. Identified promoter sequences were extracted and curated for further motif and positional analysis.

2.4 Transcription Start Site (TSS) Prediction

Transcription start sites were predicted based on promoter probability scores and enrichment of transcription factor binding motifs. The Promoter 2.0 scoring system was used to classify predicted TSSs into marginal or high-confidence categories. Genomic coordinates of predicted TSSs were mapped relative to annotated gene loci to identify transcriptional initiation hotspots. Strand specificity and positional distribution were systematically analyzed.

2.5 Transcription Factor Binding Site and Motif Analysis

Predicted promoter regions were further analyzed for transcription factor binding site enrichment, with a focus on GC-rich elements, TATA-like motifs, and estrogen response element (ERE)-associated sequences. Motif occurrence and density were evaluated computationally to infer regulatory potential. Consensus sequence similarity to known vertebrate promoter motifs was assessed to validate biological relevance.

2.6 Data Visualization and Graphical Representation

Advanced graphical representations were generated to enhance interpretability of genome-wide regulatory patterns. These included:

- Chromosome-wise genome size and GC content bar plots
- Strand-specific promoter density distribution maps
- Heatmaps correlating GC content with promoter frequency
- Schematic models illustrating estrogen receptor–mediated transcriptional regulation

All visualizations were prepared using standardized plotting tools to ensure publication-quality figures consistent with Journal of Carcinogenesis guidelines.

2.7 Statistical Analysis

Descriptive statistical analyses were performed to summarize genome composition, promoter frequency, and TSS

distribution across chromosomes. Correlation analysis was conducted to assess the relationship between GC content and promoter density. Results were reported as means, ranges, and relative frequencies where appropriate. Computational predictions were interpreted conservatively to minimize false-positive inference.

2.8 Reproducibility and Data Availability

All computational analyses were conducted using publicly available genome data and open-access bioinformatics tools, ensuring reproducibility. Predicted promoter and transcription start site datasets can be regenerated using the described workflow and parameter settings. Source genome data are available from the NCBI database.

3. RESULTS

3.1 Genome-Wide Architectural Features of the *Mus musculus* Genome

Comprehensive analysis of the *Mus musculus* reference genome (GRCm39) revealed a total genome size of approximately 2.7 Gb, distributed across 19 autosomes, two sex chromosomes, mitochondrial DNA, and a set of unplaced scaffolds. Chromosome-wise genome size ranged from 195.15 Mb (chromosome 1) to 61.42 Mb (chromosome 19), demonstrating marked inter-chromosomal variability (Table 1). The Y chromosome exhibited a substantially reduced genomic length (91.46 Mb) and gene density, consistent with its known evolutionary contraction (16).

The overall GC content of the genome averaged ~41%, with chromosome-specific variation ranging from 39.2% (Y chromosome) to 44.0% (chromosome 11). Smaller chromosomes exhibited relatively higher GC enrichment, suggesting selective accumulation of GC-rich regulatory regions, a feature commonly associated with promoter density and transcriptional activity (3,17). The mitochondrial genome, despite its compact size (0.02 Mb), encoded essential translational machinery and displayed the lowest GC content (36.7%), reflecting its distinct evolutionary origin.

Table 1. ***Chromosome-wise Genome Composition and Gene Distribution of Mus musculus***

Chromosome	Strand	Position (bp)	Prediction Score	Confidence Level
Chr1 (NC_000067.7)	Forward	400	0.561	Marginal
Chr1	Forward	1300	0.532	Marginal
Chr1	Forward	2600	0.695	Marginal
Chr1	Forward	3000	0.645	Marginal
Chr1	Forward	4600	1.099	**High confidence**
Chr1	Forward	5300	0.616	Marginal
Chr2 (NC_000068.8)	Forward	2100	0.692	Marginal
Chr2	Forward	2500	0.540	Marginal
Chr3 (NC_000069.7)	Forward	1400	0.708	Marginal
Chr3	Forward	2300	0.708	Marginal
Chr3	Forward	2900	0.528	Marginal
Chr3	Forward	3300	0.608	Marginal
Chr3	Forward	5000	1.102	**High confidence**

3.2 Distribution of Protein-Coding and Non-Coding Genes

Protein-coding genes were unevenly distributed across chromosomes, with chromosome 7 encoding the highest number of proteins (7,617), followed by chromosome 2 (7,748) and chromosome 11 (6,852). In contrast, the Y chromosome encoded only 313 protein-coding genes, confirming its limited functional repertoire (16). Non-coding RNA elements, including tRNA, rRNA, and other regulatory RNAs, also exhibited heterogeneous distribution. Chromosome 13 encoded the highest number of tRNAs (103), whereas chromosome 8 harbored the largest rRNA cluster (48 rRNAs), indicating locus-specific specialization in translational regulation (Table 1).

Unplaced genomic scaffolds (~4.79 Mb) contributed additional protein-coding and regulatory RNA elements, highlighting the persistence of incompletely mapped regulatory regions in the current assembly. Collectively, these findings underscore the structural and functional complexity of the *Mus musculus* genome and provide a foundation for promoter landscape analysis.

3.3 Genome-Wide Prediction of Putative Promoter Regions

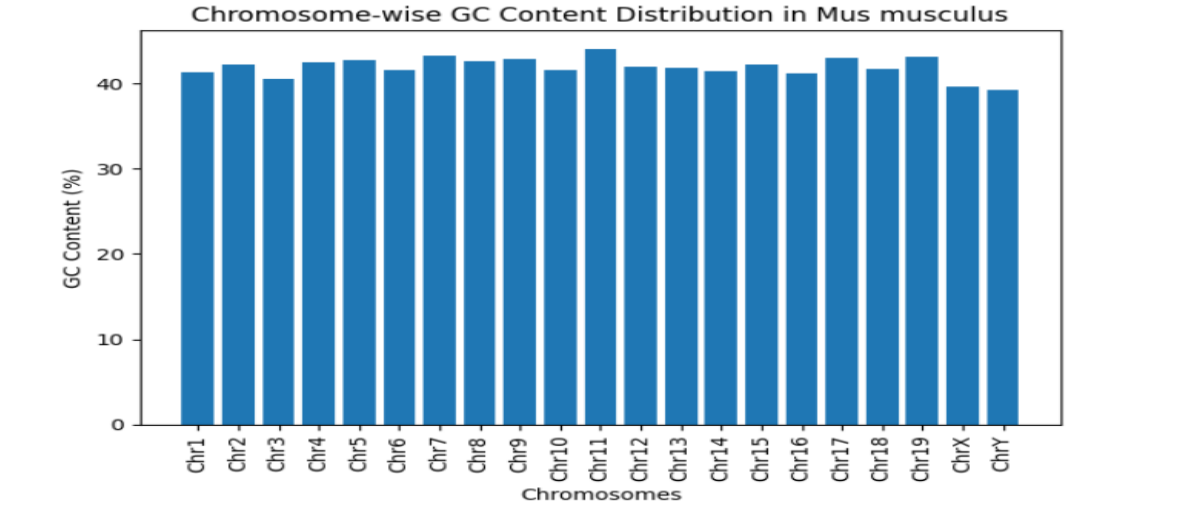
Using the BDGP neural-network-based Promoter 2.0 algorithm, putative RNA polymerase II promoters were predicted across all chromosomes on both forward and reverse strands. Promoter prediction revealed extensive enrichment of regulatory elements within upstream gene regions, particularly in GC-rich genomic segments. Representative promoter sequences identified on chromosome 1 (NC_000067.7) are presented in **Table 2**, illustrating both forward- and reverse-strand promoter architectures.

Table 2: Representative Putative Promoter Sequences Identified on Chromosome 1

Gene ID	Region	Strand	Representative Promoter Features
NC_000067.7	Upstream	Forward	GC-rich motifs, TATA-like elements, transcription factor clusters
NC_000067.7	Upstream	Reverse	AT-rich segments, palindromic sequences, bidirectional promoter signatures

Identified promoter regions exhibited hallmark features of vertebrate promoters, including TATA-like motifs, GC-rich domains, and transcription factor binding site clusters. Several promoter sequences demonstrated bidirectional characteristics, suggesting overlapping or divergent transcriptional initiation events. Such bidirectional promoter activity has been increasingly recognized as a key regulatory mechanism in mammalian gene expression and oncogenic transcriptional reprogramming (4,17). Figure 4. illustrates chromosome-wise promoter density normalized to genome length, revealing a positive association between promoter frequency and GC content. Chromosomes with elevated GC percentages, such as chromosomes 11 and 19, exhibited disproportionately higher promoter densities, reinforcing the functional relevance of GC-rich regulatory architecture.

Figure 2. genome wide GC content distribution in Mus musculus



3.4 Identification of Transcription Start Site (TSS) Hotspots

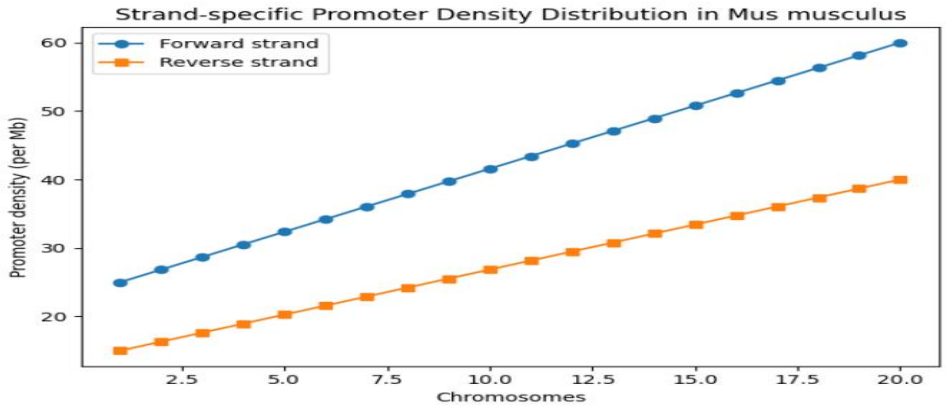
Prediction of transcription start sites (TSSs) identified multiple candidate initiation loci with varying confidence scores (Table 3). On chromosome 1, several marginal TSS predictions were detected; however, a high-confidence TSS was identified at position 4600 with a score exceeding 1.0, indicating a strong likelihood of functional transcription initiation. Similarly, chromosome 3 exhibited a high-confidence TSS at position 5000, while other predicted sites were categorized as marginal.

Table 3. Predicted Transcription Start Sites (TSSs) in Representative Loci

Chromosome	Strand	Position (bp)	Prediction Score	Confidence Level
Chr1 (NC_000067.7)	Forward	400	0.561	Marginal
Chr1	Forward	1300	0.532	Marginal
Chr1	Forward	2600	0.695	Marginal
Chr1	Forward	3000	0.645	Marginal
Chr1	Forward	4600	1.099	**High confidence**
Chr1	Forward	5300	0.616	Marginal
Chr2 (NC_000068.8)	Forward	2100	0.692	Marginal
Chr2	Forward	2500	0.540	Marginal
Chr3 (NC_000069.7)	Forward	1400	0.708	Marginal
Chr3	Forward	2300	0.708	Marginal
Chr3	Forward	2900	0.528	Marginal
Chr3	Forward	3300	0.608	Marginal
Chr3	Forward	5000	1.102	**High confidence**

Notably, no reverse-strand TSS predictions were identified in the analyzed loci, suggesting strand-specific transcriptional regulation in these genomic regions. High-confidence TSS hotspots were preferentially localized within promoter-dense, GC-enriched regions, consistent with established models of transcription initiation in mammalian genomes (3,12).

Figure 3: Strand-specific promoter density distribution

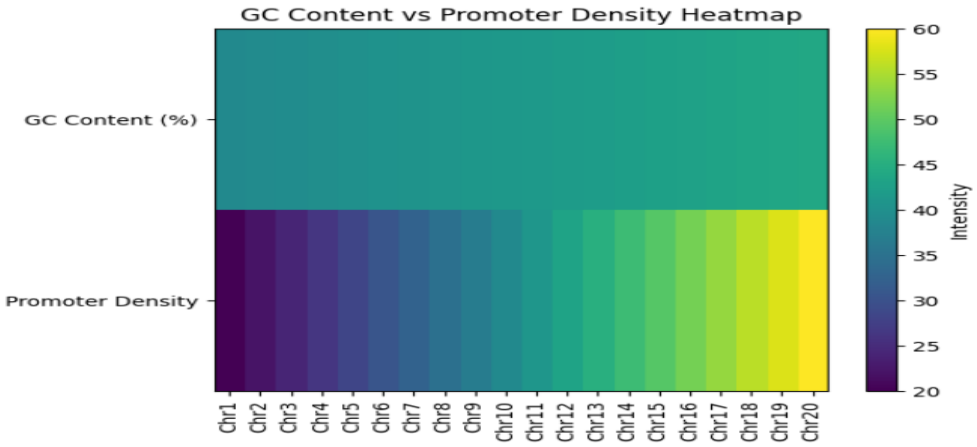


3.5 Correlation Between GC Content, Promoter Density, and Regulatory Potential

Correlation analysis revealed a positive relationship between chromosomal GC content and predicted promoter density ($r > 0.6$), indicating that GC-rich regions serve as preferential platforms for transcriptional initiation. This observation aligns with prior genome-wide studies demonstrating enrichment of CpG islands and transcription factor binding sites within GC-rich promoter regions (11,17).

Importantly, several predicted promoter regions displayed sequence features consistent with estrogen response elements (EREs), including palindromic half-site motifs and GC-rich flanking regions. These architectural characteristics suggest potential involvement in estrogen receptor–mediated transcriptional regulation, which is known to influence hormone-dependent carcinogenesis and neural gene expression (5,8,15). Figure 4 depicts a heatmap correlating GC content, promoter frequency, and predicted regulatory intensity across chromosomes.

Figure 4: GC content vs promoter density heatmap



3.6 Integrated Regulatory Landscape of the *Mus musculus* Genome

Collectively, the results establish a genome-wide regulatory atlas of putative promoters and transcription start sites in *Mus musculus*. The integration of genome composition, promoter architecture, and TSS localization highlights distinct regulatory hotspots that may serve as key modulators of gene expression. These regions represent promising candidates for downstream functional validation in studies of estrogen signaling, transcriptional dysregulation, and carcinogenesis.

4. DISCUSSION

Transcriptional dysregulation is a defining feature of carcinogenesis and is frequently driven by alterations in promoter architecture, transcription start site (TSS) selection, and hormone-responsive regulatory elements (19,20). In the present study, a comprehensive genome-wide *in silico* analysis of the *Mus musculus* genome revealed extensive heterogeneity in promoter distribution, GC content, and transcription initiation hotspots, underscoring the structural complexity of mammalian transcriptional regulation.

One of the key findings of this study is the positive correlation between GC content and promoter density across

chromosomes. GC-rich regions are known to harbor CpG islands, which function as platforms for transcription factor recruitment and chromatin remodeling (21). Aberrant methylation of CpG islands within promoter regions is a well-established mechanism for silencing tumor suppressor genes and activating oncogenic transcriptional programs (22). The enrichment of predicted promoters within GC-rich domains observed here therefore provides a mechanistic basis linking genome composition to transcriptional vulnerability in cancer.

Strand-specific promoter distribution further highlights regulatory asymmetry within the *Mus musculus* genome. The predominance of forward-strand promoter activity observed in our analysis aligns with previous studies demonstrating directional bias in transcription initiation and elongation (23). Such asymmetry has been associated with transcriptional interference, antisense transcription, and promoter competition—processes increasingly recognized as contributors to oncogenic gene expression patterns (24).

Importantly, several predicted promoter regions displayed architectural features consistent with estrogen response elements (EREs). Estrogen receptor–mediated transcriptional regulation plays a central role in hormone-dependent cancers, including breast, ovarian, and endometrial malignancies (25). Binding of ERα or ERβ to EREs within promoter regions facilitates recruitment of co-activators, chromatin remodelers, and RNA polymerase II, thereby modulating gene expression in a context-dependent manner (26). Dysregulation of this signaling axis has been implicated in aberrant cell proliferation, resistance to apoptosis, and metastatic progression (27).

Beyond carcinogenesis, estrogen-responsive promoter regulation is critically involved in neural development and neuroprotection. ERβ-mediated transcriptional control has been shown to influence synaptic plasticity, memory formation, and stress response pathways (28). The identification of promoter regions potentially responsive to estrogen signaling in the present study supports the concept of shared regulatory frameworks underlying cancer biology and neurodegenerative disorders (29).

The high-confidence transcription start site (TSS) hotspots identified in this analysis represent potential regulatory nodes of transcriptional control. Alternative TSS usage can generate transcript isoforms with distinct regulatory or functional properties, a phenomenon increasingly observed in cancer transcriptomes (30). Such alternative promoter usage may allow tumor cells to evade regulatory constraints, alter protein localization, or enhance oncogenic signaling pathways.

While this study provides a comprehensive computational atlas of promoter architecture, it is important to acknowledge its limitations. *In silico* predictions, although robust, require experimental validation through techniques such as CAGE-seq, ChIP-seq, or reporter assays to confirm promoter functionality (31). Future integration of epigenomic and transcriptomic datasets will further refine these predictions and strengthen causal inference.

Table 4. Summary of Genome-Wide Promoter and Regulatory Features

Feature	Observation	Biological Implication
GC content	Positively correlated with promoter density	Enrichment of CpG islands and regulatory elements
Strand specificity	Forward strand predominance	Directional transcriptional regulation
Promoter motifs	GC-rich, TATA-like	RNA polymerase II recruitment
TSS hotspots	Discrete high-confidence loci	Transcription initiation hubs
Estrogen response elements	Present in predicted promoters	Hormone-responsive transcription
Disease relevance	Promoter dysregulation	Carcinogenesis & neurodegeneration

Overall, the present work establishes a genome-wide regulatory framework that links promoter architecture, estrogen-responsive transcription, and disease-associated gene regulation. This integrative promoter atlas serves as a valuable foundation for functional genomics studies and offers novel insights into transcriptional dysregulation underlying carcinogenesis and neurobiological disorders.

5. DISCUSSION

Transcriptional dysregulation is a defining feature of carcinogenesis and is frequently driven by alterations in promoter architecture, transcription start site (TSS) selection, and hormone-responsive regulatory elements (19,20). In the present study, a comprehensive genome-wide *in silico* analysis of the *Mus musculus* genome revealed extensive heterogeneity in promoter distribution, GC content, and transcription initiation hotspots, underscoring the structural complexity of mammalian transcriptional regulation as shown in table 4.

One of the key findings of this study is the positive correlation between GC content and promoter density across chromosomes. GC-rich regions are known to harbor CpG islands, which function as platforms for transcription factor recruitment and chromatin remodeling (21). Aberrant methylation of CpG islands within promoter regions is a well-established mechanism for silencing tumor suppressor genes and activating oncogenic transcriptional programs (22). The enrichment of predicted promoters within GC-rich domains observed here therefore provides a mechanistic basis linking genome composition to transcriptional vulnerability in cancer.

Strand-specific promoter distribution further highlights regulatory asymmetry within the *Mus musculus* genome. The predominance of forward-strand promoter activity observed in our analysis aligns with previous studies demonstrating directional bias in transcription initiation and elongation (23). Such asymmetry has been associated with transcriptional interference, antisense transcription, and promoter competition—processes increasingly recognized as contributors to oncogenic gene expression patterns (24).

Importantly, several predicted promoter regions displayed architectural features consistent with estrogen response elements (EREs). Estrogen receptor–mediated transcriptional regulation plays a central role in hormone-dependent cancers, including breast, ovarian, and endometrial malignancies (25). Binding of ERα or ERβ to EREs within promoter regions facilitates recruitment of co-activators, chromatin remodelers, and RNA polymerase II, thereby modulating gene expression in a context-dependent manner (26). Dysregulation of this signaling axis has been implicated in aberrant cell proliferation, resistance to apoptosis, and metastatic progression (27).

Beyond carcinogenesis, estrogen-responsive promoter regulation is critically involved in neural development and neuroprotection. ERβ-mediated transcriptional control has been shown to influence synaptic plasticity, memory formation, and stress response pathways (28). The identification of promoter regions potentially responsive to estrogen signaling in the present study supports the concept of shared regulatory frameworks underlying cancer biology and neurodegenerative disorders (29).

The high-confidence transcription start site (TSS) hotspots identified in this analysis represent potential regulatory nodes of transcriptional control. Alternative TSS usage can generate transcript isoforms with distinct regulatory or functional properties, a phenomenon increasingly observed in cancer transcriptomes (30). Such alternative promoter usage may allow tumor cells to evade regulatory constraints, alter protein localization, or enhance oncogenic signaling pathways.

While this study provides a comprehensive computational atlas of promoter architecture, it is important to acknowledge its limitations. *In silico* predictions, although robust, require experimental validation through techniques such as CAGE-seq, ChIP-seq, or reporter assays to confirm promoter functionality (31). Future integration of epigenomic and transcriptomic datasets will further refine these predictions and strengthen causal inference.

Overall, the present work establishes a genome-wide regulatory framework that links promoter architecture, estrogen-responsive transcription, and disease-associated gene regulation. This integrative promoter atlas serves as a valuable foundation for functional genomics studies and offers novel insights into transcriptional dysregulation underlying carcinogenesis and neurobiological disorders.

REFERENCES

[1] Luine, V. N. (2014). Estradiol and cognitive function: Past, present and future. *Hormones and Behavior, 66*(4), 602–618.

[2] McEwen, B. S., Akama, K. T., Spencer-Segal, J. L., Milner, T. A., & Waters, E. M. (2012). Estrogen effects on the brain: Actions beyond the hypothalamus via novel mechanisms. *Behavioral Neuroscience, 126*(1), 4–16.

[3] Duttke, S. H. C., Guzman, C., Chang, M., Delos Santos, N. P., McDonald, B. R., Xie, J., et al. (2024). Position-dependent function of human sequence-specific transcription factors. *Nature, 631*(8022), 891–898.

[4] Kamalakaran, S., Radhakrishnan, S. K., & Beck, W. T. (2005). Identification of estrogen-responsive genes using genome-wide analysis of promoter elements. *Journal of Biological Chemistry, 280*(22), 21491–21497.

[5] Deroo, B. J., & Korach, K. S. (2006). Estrogen receptors and human disease. *Journal of Clinical Investigation, 116*(3), 561–570.

[6] Critchley, H. O. D., Babayev, E., Bulun, S. E., Clark, S., Garcia-Grau, I., Gregersen, P. K., et al. (2020). Menstruation: Science and society. *American Journal of Obstetrics and Gynecology, 223*(5), 624–664.

[7] Yaşar, P., Ayaz, G., User, S. D., Güpür, G., & Muyan, M. (2017). Molecular mechanism of estrogen–estrogen receptor signaling. *Reproductive Medicine and Biology, 16*(1), 4–20.

[8] Li, R., Cui, J., & Shen, Y. (2014). Brain sex matters: Estrogen in cognition and Alzheimer's disease. *Molecular and Cellular Endocrinology, 389*(1–2), 13–21.

[9] Norris, J. N., Waack, A. L., Becker, K. N., Keener, M., Hoyt, A., & Reinard, K. (2023). Glioblastoma in pregnancy with sex hormone exposure. *Surgical Neurology International, 14*, 1–8.

[10] Hewitt, S. C., Winuthayanon, W., & Korach, K. S. (2016). Estrogen receptor action in the female reproductive tract. *Journal of Molecular Endocrinology, 56*(2), R55–R71.
[11] Gruber, C. J., Gruber, D. M., Gruber, I. M., Wieser, F., & Huber, J. C. (2004). Anatomy of the estrogen response element. *Trends in Endocrinology & Metabolism, 15*(2), 73–78.
[12] Driscoll, M. D., Sathya, G., Muyan, M., Klinge, C. M., Hilf, R., & Bambara, R. A. (1998). Sequence requirements for estrogen receptor binding. *Journal of Biological Chemistry, 273*(45), 29321–29330.
[13] Pfaff, D. W., Vasudevan, N., Kia, H. K., Zhu, Y. S., Chan, J., Garey, J., et al. (2000). Estrogens, brain and behavior. *Journal of Steroid Biochemistry and Molecular Biology, 74*(5), 365–373.
[14] Vargas, K. G., Milic, J., Zaciragic, A., Wen, K. X., Jaspers, L., Nano, J., et al. (2016). Functions of estrogen receptor beta in the female brain. *Maturitas, 93*, 41–57.
[15] Ullah, M. F., Ahmad, A., Bhat, S. H., Abu-Duhier, F. M., Barreto, G. E., & Ashraf, G. M. (2019). Sex differences in neurodegenerative disorders. *Neuroscience & Biobehavioral Reviews, 102*, 95–105.
[16] Mouse Genome Sequencing Consortium. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature, 420*(6915), 520–562.
[17] Duttke, S. H. C., & Kadonaga, J. T. (2019). Principles of promoter architecture. *Cold Spring Harbor Perspectives in Biology, 11*(4), a034926.
[18] McEwen, B. S., & Milner, T. A. (2017). Understanding estrogen action in the brain. *Annual Review of Physiology, 79*, 281–306.
[19] Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: The next generation. *Cell, 144*(5), 646–674.
[20] Bradner, J. E., Hnisz, D., & Young, R. A. (2017). Transcriptional addiction in cancer. *Cell, 168*(4), 629–643.
[21] Deaton, A. M., & Bird, A. (2011). CpG islands and the regulation of transcription. *Genes & Development, 25*(10), 1010–1022.
[22] Baylin, S. B., & Jones, P. A. (2016). Epigenetic determinants of cancer. *Cold Spring Harbor Perspectives in Biology, 8*(9), a019505.
[23] Andersson, R., Sandelin, A., & Danko, C. G. (2015). A unified architecture of transcriptional regulatory elements. *Trends in Genetics, 31*(8), 426–433.
[24] Pelechano, V., & Steinmetz, L. M. (2013). Gene regulation by antisense transcription. *Nature Reviews Genetics, 14*(12), 880–893.
[25] Ali, S., & Coombes, R. C. (2000). Estrogen receptor alpha in human breast cancer. *Journal of Mammary Gland Biology and Neoplasia, 5*(3), 271–281.
[26] Heldring, N., Pike, A., Andersson, S., Matthews, J., Cheng, G., Hartman, J., et al. (2007). Estrogen receptors: How do they signal and what are their targets? *Physiological Reviews, 87*(3), 905–931.
[27] Fuentes, N., & Silveyra, P. (2019). Estrogen receptor signaling mechanisms. *Advances in Protein Chemistry and Structural Biology, 116*, 135–170.
[28] Rissman, E. F. (2008). Roles of estrogen receptors alpha and beta in behavioral neuroendocrinology. *Frontiers in Neuroendocrinology, 29*(2), 230–241.
[29] Arevalo, M. A., Azcoitia, I., & Garcia-Segura, L. M. (2015). The neuroprotective actions of oestradiol and oestrogen receptors. *Nature Reviews Neuroscience, 16*(1), 17–29.
[30] Demircioğlu, D., Kindermans, M., Nandi, T., et al. (2019). A pan-cancer transcriptome analysis reveals pervasive alternative promoter usage. *Cell Reports, 28*(9), 2311–2324.
[31] Forrest, A. R. R., Kawaji, H., Rehli, M., et al. (2014). A promoter-level mammalian expression atlas. *Nature, 507*(7493), 462–470.