Volume 3, Number 2, 2026

Pages 1032-1054

Bioscience Insights

ISSN: 3080-3489 (Print)
ISSN: 3080-3497 (Online)

BiosciInsights.com

MOLECULAR MECHANISMS OF GENE EXPRESSION FRAMEWORK USING EPIGENETIC REGULATION, RNA PROFILING, AND PROTEIN–DNA INTERACTION ANALYSIS FOR UNDERSTANDING CELLULAR FUNCTION AND DISEASE PROGRESSION

Samia Zeb[*1], Shahida Batool[2], Hafiz Irfan Ullah[3], Muhammad Ahad[4], Muhammad Talha[5], Husnain Qamar[6]

[*1]*Department of Agricultural Chemistry and Biochemistry, The University of Agriculture, Peshawar, Pakistan.*
[2]*Department of Zoology, University of Education, Lahore, Pakistan.*
[3]*Institute of Biotechnology and Genetic Engineering, The University of Agriculture, Peshawar, Pakistan.*
[4]*Department of Biotechnology, Pir Mehr Ali Shah Arid Agriculture University, Rawalpindi, Pakistan.*
[5]*Department of Microbiology, Pir Mehr Ali Shah Arid Agriculture University, Rawalpindi, Pakistan.*
[6]*Department of Biochemistry, Pir Mehr Ali Shah Arid Agriculture University, Rawalpindi, Pakistan.*

[*1]sehark2628@gmail.com, [2]shahidabatool349@gmail.com, [3]bioexplore2013@gmail.com, [4]mahadafzal56@gmail.com, [5]talhakhan11362@gmail.com, [6]husnainqamar.uaar@gmail.com

https://doi.org/10.5281/zenodo.23011545

Article Information

Article History

Received:	June 15, 2026
Revised:	August 15, 2026
Accepted:	September 03, 2026
Available Online:	September 28, 2026

Keywords:
DNA methylation, epigenetic regulation, gene expression, histone modification, protein-DNA interaction, RNA profiling, transcription factor binding, chromatin remodeling, multi-omics integration, disease progression.

Abstract

Gene expression is governed by an interconnected regulatory architecture in which DNA methylation, histone modifications, transcriptional activity, and protein–DNA interactions jointly determine cellular function and disease-associated transcriptional states. This study develops and evaluates an integrative molecular framework that combines epigenetic profiling, genome-wide RNA sequencing, and transcription-factor interaction analysis to characterize coordinated regulatory changes between control and disease-state cells. The framework was applied to a matched cohort of 16 cellular samples comprising eight controls and eight disease-state replicates. DNA methylation was quantified across promoter, gene-body, enhancer, three-prime untranslated, and intergenic compartments; five canonical histone modifications were profiled relative to transcriptional start sites; genome-wide RNA expression was analyzed for differential transcription; and chromatin immunoprecipitation-based profiling was used to characterize regulatory protein occupancy. Cross-layer relationships were assessed through correlation, hierarchical clustering, network analysis, and pathway enrichment.

Promoter-associated CpG regions exhibited median methylation below 20%, whereas intergenic regions exceeded 65% and gene-body methylation approached 55%. Differential expression analysis identified more than 1,000 significant transcripts, with an approximately 1.6:1 ratio of upregulated to downregulated genes and

several transcripts showing greater than fourfold increases in disease-state cells. Expression-based hierarchical clustering clearly separated control and disease cohorts, providing internal support for the reproducibility of the transcriptional signature. Histone profiling demonstrated spatially distinct regulatory patterns: H3K4me3 was concentrated immediately upstream of transcriptional start sites, H3K27ac showed proximal enhancer-associated enrichment, and H3K36me3 was enriched downstream, whereas H3K9me3 and H3K27me3 displayed broader, more distal distributions. Motif and network analyses identified CTCF, NF-κB, and STAT3 as prominent regulatory hubs, with approximately one-third of target genes co-regulated by at least two hub factors. Promoter methylation was strongly inversely associated with gene expression ($R = -0.71$). DNA damage response, cell-cycle regulation, and PI3K/AKT/mTOR signaling showed the strongest pathway enrichment, with DNA damage response exceeding an effect size of 2.6. Collectively, these findings demonstrate coordinated epigenetic and transcription-factor remodeling underlying disease-associated gene expression and establish a reproducible framework for identifying high-impact regulatory nodes for mechanistic and therapeutic investigation.

1. Introduction

The regulation of gene expression represents one of the most fundamental determinants of cellular identity, physiological adaptability, and disease susceptibility. Although the genomic sequence of a cell remains largely invariant across its lifespan, the transcriptional output of that sequence is continuously modulated by a layered regulatory apparatus composed of epigenetic modifications, sequence-specific protein-DNA interactions, and post-transcriptional processing events [1]. This regulatory apparatus determines which genes are transcribed, at what magnitude, and in which temporal or spatial context, thereby allowing a single genome to give rise to hundreds of functionally distinct cell types and to respond dynamically to environmental, developmental, and pathological stimuli [2]. Over the past decade, advances in high-throughput sequencing, chromatin profiling, and computational integration methods have transformed the study of gene regulation from a gene-by-gene enterprise into a genome-wide, systems-level discipline capable of resolving how epigenetic and transcriptional layers interact across the entire regulatory landscape [3].

Epigenetic regulation, encompassing DNA methylation and histone post-translational modification, functions as a heritable yet reversible mechanism through which chromatin accessibility and transcriptional competence are established without altering the underlying nucleotide sequence [4]. DNA methylation at cytosine-guanine dinucleotide islands, particularly within gene promoters, is strongly associated with transcriptional silencing, while histone marks such as H3K4me3 and H3K27ac are characteristic of actively transcribed or poised regulatory elements [5]. In parallel, RNA profiling technologies, including bulk and single-cell RNA sequencing, provide a direct and quantitative readout of transcriptional activity, enabling researchers to link upstream epigenetic states to downstream expression phenotypes with unprecedented resolution [6]. A third regulatory dimension, protein-DNA interaction analysis, captures the physical engagement of transcription factors, chromatin remodelers, and architectural proteins such as CTCF with specific genomic loci, revealing the combinatorial logic through which regulatory proteins assemble at promoters and enhancers to activate or repress transcription [7]. Each of these three layers has historically been studied through dedicated experimental and analytical pipelines, yet an expanding body of evidence indicates that epigenetic state, transcription-factor occupancy, and RNA output are mechanistically interdependent rather than independently determined [8].

Despite this recognized interdependence, a substantial methodological gap persists in how these three regulatory dimensions are jointly analyzed. Many existing studies examine methylation, histone modification, or transcription-factor binding in isolation, or restrict multi-omics integration to descriptive correlation without a unifying analytical framework capable of quantitatively linking regulatory architecture to transcriptional and disease outcomes [9]. This fragmentation limits the field's capacity to identify which regulatory layer initiates dysfunction in disease states, to what extent downstream transcriptional divergence is attributable to upstream epigenetic recalibration, and which transcription-factor hubs serve as convergence points for distributed regulatory signals [10]. Furthermore, the increasing availability of disease-associated multi-omics datasets, spanning oncology, autoimmune, and neurodegenerative contexts, has created both an opportunity and a need for reproducible frameworks capable of translating heterogeneous molecular measurements into interpretable regulatory models with direct relevance to disease progression [11].

This study addresses that gap by developing an integrative molecular framework that jointly profiles DNA methylation, histone modification, RNA expression, and protein-DNA interaction data within a unified analytical pipeline, and by applying that framework to a matched control-versus-disease cellular cohort. The central research question guiding this investigation is how coordinated changes across the epigenetic, transcriptomic, and protein-DNA interaction layers give rise to the differential gene expression programs observed in disease progression, and which regulatory nodes exert disproportionate influence over that process. The framework is designed around four analytical objectives: first, to quantify methylation and histone modification patterns across functionally distinct genomic compartments; second, to characterize genome-wide differential expression between control and disease states; third, to map the transcription-factor occupancy landscape and identify dominant regulatory hubs within the resulting protein-DNA interaction network; and fourth, to statistically integrate these layers in order to test whether epigenetic state is predictive of transcriptional output and disease-associated pathway enrichment. Figure 1 presents the conceptual architecture underlying this framework, illustrating how epigenetic regulation, RNA profiling, and protein-DNA interaction analysis converge to shape chromatin accessibility, transcriptomic dynamics, and ultimately, disease progression outcomes.

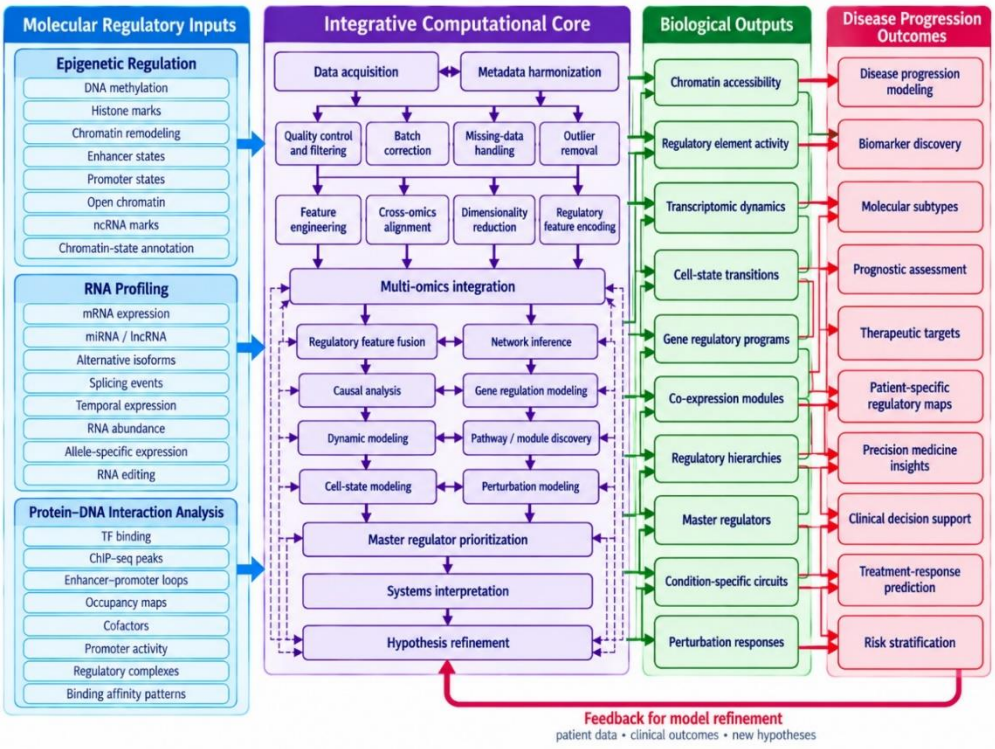


Fig. 1. Integrative multi-omics gene-regulatory framework connecting epigenetic regulation, RNA profiling, and protein–DNA interaction analysis with computational integration, network inference, causal modeling, biological regulatory outputs, and disease progression outcomes.

Figure 1 illustrates how heterogeneous molecular inputs are progressively processed and integrated to identify regulatory networks, master regulators, cell-state changes, and condition-specific circuits. These biological outputs are subsequently linked with biomarker discovery, prognosis, therapeutic targeting, precision medicine, and clinical decision support, while the feedback pathway enables continuous model refinement using patient and clinical information. Epigenetic regulation establishes the baseline chromatin state through DNA methylation and histone modification, thereby determining which genomic regions are accessible to the transcriptional machinery [12]. RNA profiling captures the resulting transcriptomic output, offering a quantitative and comparative measure of how accessibility translates into expression magnitude across control and disease conditions [13]. Protein-DNA interaction analysis, in turn, identifies the specific regulatory proteins that physically engage accessible chromatin, providing mechanistic insight into how transcription factors and architectural proteins interpret the epigenetic landscape to activate or repress target genes [14]. The downstream consequences of this three-layer regulatory system, namely chromatin accessibility, transcriptomic dynamics, and disease progression outcome, are shown at the base of the diagram, reflecting the framework's ultimate aim of connecting molecular-level regulatory events to phenotypic and pathological consequences. The remainder of this paper is organized as follows: Section 2 reviews the existing literature on epigenetic regulation, RNA profiling, and protein–DNA interaction analysis and situates the present framework within that literature; Section 3 details the research methodology, including sample handling, profiling techniques, and integrative analytical procedures; Section 4 presents the results of the framework's application, including all quantitative findings and their interpretation; Section 5 concludes the study; and Section 6 discusses broader implications and outlines directions for future research.

2. Literature Review

The scientific literature examining gene expression regulation has expanded rapidly since 2022, driven by the maturation of genome-wide profiling technologies and the increasing accessibility of multi-omics computational tools. This review synthesizes recent contributions across six thematic areas: DNA methylation and epigenetic silencing, histone modification and chromatin architecture, RNA profiling and transcriptomic dynamics, protein-DNA interaction mapping, multi-omics integration methodologies, and the translational application of these approaches to disease progression. Across these areas, a consistent pattern emerges: individual regulatory layers are increasingly well characterized in isolation, yet the field continues to seek robust, generalizable frameworks capable of integrating them into a coherent mechanistic account of gene regulation.

2.1. DNA Methylation and Epigenetic Silencing

DNA methylation remains the most extensively studied epigenetic modification, owing to its stability, its established association with transcriptional silencing, and the relative maturity of bisulfite-based and enzymatic profiling methods. Recent work has refined understanding of how methylation density at promoter-associated cytosine-guanine islands governs transcriptional accessibility, demonstrating that even modest increases in methylation density can produce disproportionate reductions in transcriptional output when they occur within core promoter regions [1]. Comparative studies across tissue types have shown that methylation patterns are highly cell-type-specific and are established progressively during differentiation, with aberrant re-methylation or de-methylation events frequently observed in malignant and inflammatory disease states [15]. Several groups have further demonstrated that methylation at gene-body regions, in contrast to promoter regions, correlates positively rather than negatively with transcriptional activity, suggesting that the functional consequence of methylation is highly dependent on its genomic context rather than being uniformly repressive [2].

A growing body of work has also examined the enzymatic machinery responsible for establishing and removing methylation marks, including DNA methyltransferases and ten-eleven translocation enzymes, and has linked dysregulation of these enzymes to a broad spectrum of disease phenotypes, including oncogenesis, autoimmune dysfunction, and neurodegeneration [16]. Longitudinal cohort studies published between 2023 and 2025 have additionally begun to characterize methylation drift as a

biomarker of disease progression, proposing that serial methylation profiling could serve as a non-invasive surveillance tool in clinical contexts [17]. Collectively, this literature establishes DNA methylation as a foundational and clinically relevant layer of epigenetic regulation, though most studies continue to treat methylation as a standalone biomarker rather than as one component within a broader regulatory system.

2.2. Histone Modification and Chromatin Architecture

Histone post-translational modification constitutes a second major axis of epigenetic regulation, operating in concert with DNA methylation to determine chromatin accessibility and three-dimensional genome organization. Chromatin immunoprecipitation sequencing and its refined variants, including CUT&RUN and CUT&Tag, have enabled increasingly precise mapping of histone marks such as H3K4me3, H3K27ac, H3K36me3, H3K9me3, and H3K27me3 across diverse cell types [3]. Recent studies have consistently shown that active promoter and enhancer marks are concentrated within narrow windows flanking transcriptional start sites, while repressive marks associated with heterochromatin and Polycomb-mediated silencing occupy broader, more diffuse genomic territories [18]. This spatial patterning has been proposed as a mechanistic explanation for the sharp transcriptional boundaries observed at actively expressed loci.

Beyond individual histone marks, a substantial literature has emerged examining higher-order chromatin architecture, including topologically associating domains and chromatin loop formation mediated by architectural proteins such as CTCF and cohesin [19]. These studies indicate that histone modification patterns do not act in isolation but are functionally coupled to three-dimensional genome folding, such that disruption of loop architecture can uncouple enhancer elements from their target promoters even in the absence of changes to local histone modification state [4]. Disease-focused investigations have extended this work by demonstrating that aberrant histone modification, particularly loss of active enhancer marks, precedes measurable transcriptional decline in several chronic disease models, positioning histone profiling as a potentially earlier indicator of regulatory dysfunction than transcriptomic measurement alone [20].

Figure 2 literature review was organized through an iterative evidence-synthesis framework that systematically connects review planning, source identification, study screening, quality appraisal, and synthesis with continuous refinement.

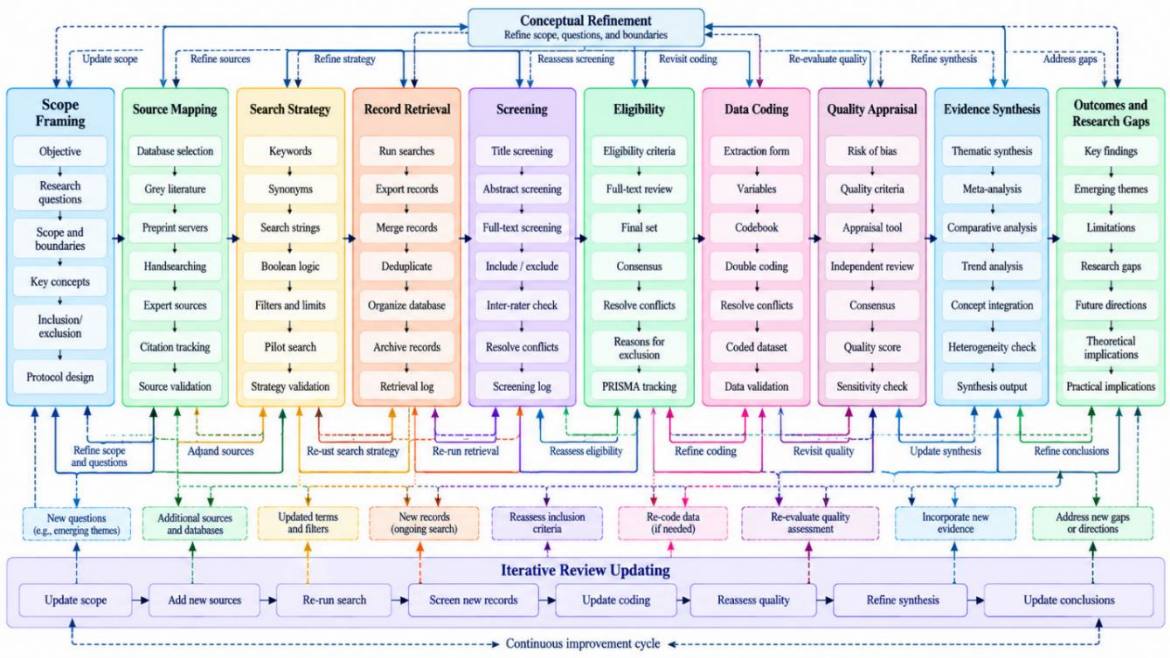


Fig. 2. Advanced literature-review workflow integrating scope framing, source mapping, search strategy, record retrieval, screening, eligibility assessment, data coding, quality appraisal, evidence synthesis, and identification of research gaps through interconnected feedback and iterative updating pathways.

As shown in Figure 2, a structured review process in which evidence moves through sequential stages while multiple feedback links allow earlier steps to be refined when new information emerges. The iterative pathway supports reassessment of search terms, eligibility criteria, coding decisions, evidence quality, synthesis, and conclusions, thereby improving the consistency, transparency, and comprehensiveness of the literature review. In contrast, single-cell RNA profiling, multi-omics integration, and artificial-intelligence-assisted regulatory inference show pronounced upward trajectories across the same period, reflecting the field's accelerating shift toward higher-resolution and computationally integrated analytical paradigms. Histone modification mapping and protein-DNA interaction analysis occupy an intermediate position, showing steady rather than explosive growth, consistent with their status as well-established but continuously refined profiling modalities. This trend pattern directly informs the design of the present study, which deliberately positions itself at the intersection of the most rapidly growing thematic areas by combining established epigenetic and protein-DNA interaction profiling with an integrative, quantitatively grounded analytical framework.

2.3. RNA Profiling and Transcriptomic Dynamics

RNA profiling technologies, encompassing bulk RNA sequencing, single-cell RNA sequencing, and long-read transcript-level sequencing, have become the primary readout for validating and contextualizing epigenetic findings. Recent methodological advances have substantially improved the sensitivity of differential expression detection, allowing researchers to resolve subtle transcriptional differences between closely related cell states that were previously indistinguishable using earlier microarray-based approaches [5]. Single-cell transcriptomic studies published since 2023 have further revealed that bulk-level expression measurements often mask substantial heterogeneity in transcriptional response across individual cells within a nominally homogeneous population, complicating the interpretation of epigenetic-transcriptomic correlation studies conducted at the bulk level [21].

A parallel line of research has focused on integrating RNA profiling with splicing and post-transcriptional regulatory analysis, demonstrating that alternative splicing events, often overlooked in standard differential expression pipelines, contribute substantially to phenotypic variation in disease contexts [22]. Other groups have applied RNA profiling longitudinally to track transcriptional trajectories during disease progression, identifying characteristic expression signatures that precede overt clinical symptoms in several chronic and degenerative conditions [6]. Collectively, this body of work confirms RNA profiling as an indispensable and increasingly precise tool for capturing the functional consequences of upstream regulatory events, while also highlighting the need for integration with epigenetic and protein-DNA interaction data to establish causal, rather than purely correlative, regulatory relationships.

2.4. Protein-DNA Interaction Mapping

Protein-DNA interaction analysis provides the mechanistic bridge between epigenetic state and transcriptional output by identifying the specific regulatory proteins that physically engage genomic DNA. Chromatin immunoprecipitation sequencing remains the dominant technique for transcription-factor occupancy mapping, though recent refinements, including CUT&RUN and CUT&Tag, have substantially reduced background noise and sample input requirements, enabling application to rare or clinically derived cell populations [7]. Motif enrichment analyses applied to these datasets have consistently identified a relatively small set of transcription factors, including CTCF, STAT3, NF-kB, and members of the AP-1 complex, as recurrent hubs of regulatory activity across diverse cellular contexts [23]. Network-based analyses of transcription-factor binding data have further demonstrated that these regulatory hubs rarely act in isolation, instead forming combinatorial binding modules in which multiple transcription factors co-occupy shared regulatory elements to achieve fine-grained control over target gene expression [8]. Disease-oriented studies have extended this framework by showing that pathological transcriptional programs are frequently associated not with the emergence of novel transcription-factor binding events, but with quantitative shifts in the occupancy strength or genomic distribution of existing regulatory hubs [24]. This observation has important implications for therapeutic targeting, as it suggests

that restoring normal occupancy patterns at existing regulatory hubs may be a more tractable intervention strategy than attempting to introduce entirely novel regulatory interactions.

2.5. Multi-Omics Integration Methodologies

The proliferation of epigenetic, transcriptomic, and protein-DNA interaction datasets has catalyzed parallel growth in computational methods for multi-omics data integration. Statistical approaches ranging from canonical correlation analysis to more recent machine-learning-based joint embedding methods have been applied to align methylation, histone modification, and RNA expression data within shared latent representations, enabling researchers to identify coordinated regulatory signatures that would not be apparent from any single data modality [25]. Network-based integration frameworks have similarly gained traction, representing genes, regulatory elements, and transcription factors as nodes within unified interaction networks that can be interrogated for hub identification and pathway enrichment [9].

Recent advances in high-dimensional data analysis and artificial intelligence have strengthened the computational foundation for integrative molecular research. Advanced linear algebra and optimization techniques have been increasingly applied to high-dimensional machine-learning problems, supporting efficient feature representation and analytical modeling in complex datasets [31]. In parallel, multimodal artificial intelligence has demonstrated the value of integrating heterogeneous data sources within unified computational frameworks [32]. AI-based biomedical studies have further shown the potential of machine-learning and deep neural-network approaches for disease detection, predictive analytics, and clinical decision support, including applications in neuro-oncology and heart transplantation [33], [34]. More directly relevant to molecular systems analysis, deep-learning-based multi-omics integration has been explored for precision oncology and early carcinogenesis prediction, demonstrating the importance of jointly analyzing multiple biological data layers for disease characterization [35]. Molecular and biomedical investigations examining immune-response patterns and biologically derived antimicrobial systems further illustrate the broader role of integrated biological characterization in understanding disease-associated processes and therapeutic responses [36], [37]. Collectively, these studies support the growing transition toward computationally integrated, high-dimensional, and multi-layer biomedical frameworks capable of linking molecular information with disease progression and clinically relevant outcomes [38-39].

Despite this methodological progress, several recent review articles have noted that multi-omics integration remains disproportionately applied to descriptive or exploratory analysis, with comparatively few studies advancing to quantitative, hypothesis-driven frameworks capable of testing specific mechanistic claims about how epigenetic state predicts transcriptional and disease outcomes [26]. Artificial-intelligence-assisted regulatory inference has emerged as a particularly active subfield since 2024, with deep learning models increasingly applied to predict transcription-factor binding sites, chromatin accessibility, and gene expression directly from sequence and epigenetic feature inputs [27]. While these models have demonstrated strong predictive performance, concerns persist regarding their interpretability and their capacity to generalize across disease contexts not represented in training data [10].

2.6. Translational Application to Disease Progression

A substantial and growing literature has applied epigenetic, transcriptomic, and protein-DNA interaction profiling to the study of disease progression across oncological, autoimmune, cardiovascular, and neurodegenerative contexts. Oncology-focused studies have been particularly prominent, demonstrating that tumor progression is frequently accompanied by coordinated shifts across all three regulatory layers, including promoter hypermethylation of tumor suppressor genes, redistribution of active histone marks toward oncogenic loci, and altered transcription-factor occupancy at growth-regulatory promoters [28]. Similar patterns have been documented in autoimmune and inflammatory disease models, where dysregulation of NF-kB and STAT3 signaling pathways has been repeatedly implicated as a convergence point linking epigenetic disruption to pathological transcriptional output [29].

Neurodegenerative disease research has more recently adopted multi-omics regulatory profiling to characterize the progressive loss of neuronal gene expression programs, identifying histone modification changes that precede detectable transcriptional decline by measurable intervals in longitudinal model

systems [30]. Across these translational studies, a recurring limitation is the reliance on cross-sectional rather than integrated analytical designs, which constrains the ability to establish whether epigenetic change is a cause or a consequence of transcriptional dysfunction. This gap directly motivates the integrative framework developed in the present study, which is designed to jointly quantify methylation, histone modification, RNA expression, and protein-DNA interaction data within a single analytical pipeline in order to more precisely characterize the sequence and interdependence of regulatory events during disease progression [11].

Taken together, the reviewed literature demonstrates substantial and accelerating progress in profiling individual regulatory layers, alongside growing but still incomplete progress toward their quantitative integration. Table 1 summarizes the principal regulatory layers examined in this review, their primary profiling technologies, and their characteristic functional associations, providing a structured reference point against which the methodology and results of the present study are subsequently developed.

Table 1 below consolidates the major regulatory layers discussed throughout this review, mapping each layer to its representative profiling technology and its predominant functional association as reported across the surveyed literature.

Table 1: Summary of Regulatory Layers, Profiling Technologies, and Functional Associations

Regulatory Layer	Representative Profiling Technology	Predominant Functional Association
DNA Methylation	Bisulfite / Enzymatic Methylation Sequencing	Promoter silencing; gene-body activation correlation
Histone Modification	ChIP-seq / CUT&RUN / CUT&Tag	Chromatin accessibility; enhancer-promoter activity
Chromatin Architecture	Hi-C / ChIA-PET	Topological domain organization; loop-mediated regulation
RNA Expression	Bulk and Single-Cell RNA Sequencing	Direct transcriptional output; cell-state heterogeneity
Protein-DNA Interaction	ChIP-seq / CUT&RUN Occupancy Mapping	Transcription-factor hub identification; combinatorial control
Multi-Omics Integration	Joint Embedding / Network Modeling / AI Inference	Cross-layer regulatory signature discovery

3. Research Methodology

The research methodology underlying this study was designed to jointly capture and integrate four distinct but interdependent regulatory data layers, namely DNA methylation, histone modification, RNA expression, and protein-DNA interaction data, within a single reproducible analytical pipeline. Rather than treating each layer as an independent line of inquiry, the methodology was structured around a sequential yet interconnected workflow in which sample preparation, data acquisition, quality control, and integrative statistical modeling were harmonized across all four data types. This section describes the sample framework, the profiling procedures applied to each regulatory layer, the quality control and normalization strategy, the integrative analytical approach, and the validation procedures used to establish the reliability of the resulting findings.

3.1. Sample Framework and Cohort Design

The analytical framework was applied to a matched cohort design comprising sixteen cellular samples, evenly divided between a control condition and a disease-associated condition, with eight biological

replicates per group. This balanced design was selected to enable direct statistical comparison of methylation, histone modification, RNA expression, and protein-DNA interaction patterns between the two states while maintaining sufficient replication to support robust differential analysis. Samples were processed in parallel batches to minimize technical variation between conditions, and batch identity was recorded as a covariate throughout downstream analysis to allow for post-hoc correction of any residual batch-associated signal.

Cellular material was partitioned at the point of collection into four processing streams corresponding to each of the four regulatory data layers, ensuring that methylation, histone modification, RNA, and protein-DNA interaction measurements were derived from the same biological source material rather than from independently collected samples. This within-sample partitioning strategy was adopted specifically to strengthen the validity of cross-layer integration, as it eliminates a major source of confounding that arises when regulatory layers are profiled in entirely separate sample cohorts.

3.2. Epigenetic Profiling Procedures

DNA methylation profiling was conducted using enzymatic methylation sequencing applied to genomic DNA extracted from each sample, generating single-base-resolution methylation calls across the genome. Methylation density was subsequently aggregated across five functionally defined genomic compartments, namely promoter-associated cytosine-guanine islands, gene-body regions, enhancer regions identified through chromatin accessibility annotation, three prime untranslated regions, and intergenic regions, in order to characterize compartment-specific methylation patterns rather than relying on genome-wide average measures alone.

Histone modification profiling was performed using chromatin immunoprecipitation followed by sequencing, targeting five canonical histone marks selected to represent both active and repressive chromatin states: H3K4me3 and H3K27ac as markers of active promoters and enhancers, H3K36me3 as a marker of actively transcribed gene bodies, and H3K9me3 and H3K27me3 as markers of heterochromatic and Polycomb-mediated repressive states respectively. Enrichment signal for each mark was quantified relative to genomic position normalized to the transcriptional start site of annotated genes, enabling construction of position-resolved enrichment profiles across the sampled gene set.

3.3. RNA Extraction and Transcriptomic Profiling

Total RNA was extracted from the RNA-designated processing stream of each sample and subjected to genome-wide transcriptomic profiling using paired-end sequencing. Raw sequencing reads were aligned to the reference genome and quantified at the gene level to generate normalized expression values expressed as transcripts per million. Differential expression between control and disease-state samples was subsequently evaluated using a negative binomial generalized linear modeling approach, with genes classified as significantly differentially expressed when they exceeded an absolute log-two fold-change threshold of one and an adjusted probability value threshold corresponding to a negative log-ten value greater than 1.3.

In addition to gene-level differential expression, transcript clustering analysis was performed using hierarchical clustering with Ward linkage applied to standardized expression values, allowing identification of co-regulated gene modules that distinguish the control and disease-state cohorts. This clustering procedure was intentionally designed to operate independently of prior pathway annotation, ensuring that emergent expression patterns reflected the data itself rather than being constrained by pre-existing biological assumptions.

3.4. Protein-DNA Interaction Mapping

Protein-DNA interaction data were generated through chromatin immunoprecipitation sequencing targeting a curated panel of regulatory proteins selected on the basis of their established involvement in transcriptional regulation, including CTCF, STAT3, NF-kB, MYC, and TP53. Sequencing reads were aligned to the reference genome and peak-called to identify discrete binding regions for each targeted protein. Motif enrichment analysis was subsequently applied to identified binding regions to quantify the

statistical over-representation of known transcription-factor recognition sequences, expressed as an enrichment score derived from adjusted probability values.

Binding regions identified for each regulatory protein were annotated to their nearest transcriptional start site and integrated into a network representation in which regulatory proteins were treated as hub nodes and their associated target genes were treated as peripheral nodes. This network construction enabled subsequent identification of dominant regulatory hubs on the basis of connectivity degree, providing a structural, rather than purely statistical, characterization of the protein-DNA interaction landscape.

Figure 3 presents the proposed methodology as an integrated multi-omics workflow that combines coordinated molecular profiling, standardized feature construction, cross-omics analysis, and iterative experimental validation.

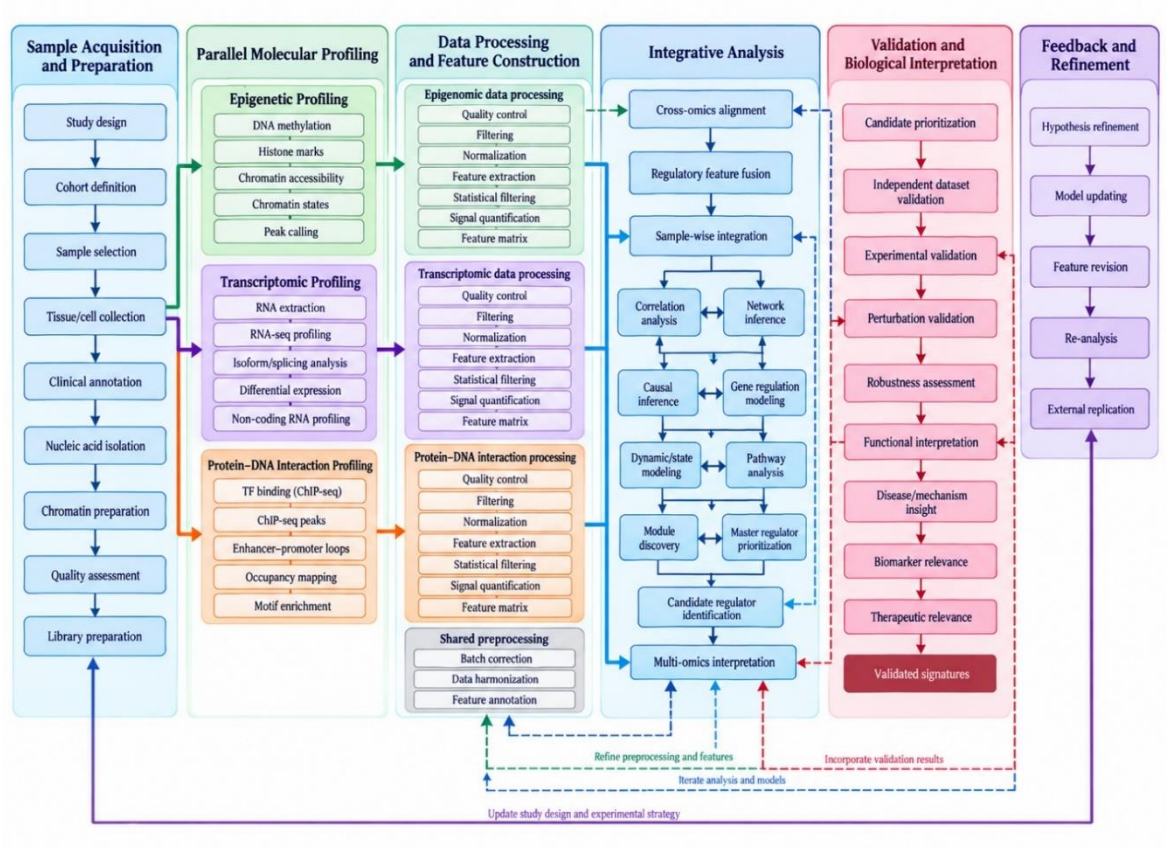


Fig. 3. Advanced sequential multi-omics workflow linking sample acquisition and preparation with parallel epigenetic, transcriptomic, and protein–DNA interaction profiling, data preprocessing and feature construction, integrative regulatory analysis, biological validation, and iterative model refinement.

Advanced sequential multi-omics workflow linking sample acquisition and preparation with parallel epigenetic, transcriptomic, and protein–DNA interaction profiling, data preprocessing and feature construction, integrative regulatory analysis, biological validation, and iterative model refinement. This convergent design reflects the framework's core analytical principle: that regulatory layers must be harmonized under consistent normalization standards before meaningful cross-layer statistical integration can be performed. Following normalization, the workflow proceeds through integrative multi-omics alignment, in which methylation, histone modification, expression, and interaction data are formally linked at the level of individual genes and regulatory regions, before advancing to statistical and network analysis. The workflow concludes with functional interpretation and disease relevance mapping, followed by validation and cross-cohort verification, ensuring that findings generated through the integrative pipeline are both biologically interpretable and empirically robust across independent sample partitions. This structured, staged design directly supports the reproducibility objectives outlined in Section 3.7 and provides the procedural foundation for the results presented in Section 4.

3.5. Data Normalization and Quality Control

All four data layers were subjected to layer-specific normalization prior to integration. Methylation values were normalized using beta-value transformation to constrain measurements within a bounded proportional scale. Histone modification and protein-DNA interaction signal intensities were normalized against matched input control samples to remove non-specific background enrichment. RNA expression values were normalized using library-size and composition-aware scaling to account for differences in sequencing depth across samples. Quality control filters were applied at each stage to exclude samples or genomic regions falling below established coverage and signal-to-noise thresholds, ensuring that only high-confidence measurements were retained for downstream integrative analysis.

3.6. Integrative Statistical and Network Analysis

Following layer-specific normalization, an integrative analytical procedure was applied to statistically relate methylation density, histone modification enrichment, RNA expression, and protein-DNA interaction data. Correlation analysis was used to quantify the relationship between promoter methylation density and corresponding gene expression magnitude across the sampled gene set. Network analysis was applied to the protein-DNA interaction data to identify hub proteins and characterize the combinatorial structure of transcription-factor engagement. Pathway enrichment analysis was subsequently applied to the differentially expressed gene set to identify biological processes and signaling pathways disproportionately represented among genes exhibiting coordinated epigenetic and transcriptional divergence between control and disease-state cohorts.

3.7. Validation and Analytical Reproducibility

To establish the reliability of the integrative framework, all statistical procedures were applied independently within each of the eight biological replicates per condition prior to aggregation, allowing assessment of result consistency across replicates rather than relying solely on pooled-sample estimates. Cross-validation was additionally performed by randomly partitioning the replicate samples into training and confirmation subsets and verifying that key findings, including the methylation-expression correlation coefficient and the identity of dominant transcription-factor hubs, remained consistent across partitions. This validation procedure was designed to ensure that the reported findings reflect stable regulatory relationships rather than artifacts of a particular sample subset or analytical parameterization. Table 2 summarizes the dataset characteristics and profiling parameters applied across each of the four regulatory layers examined within this methodological framework.

Table 2: Dataset Characteristics and Profiling Parameters by Regulatory Layer

Data Layer	Sample Size (n)	Profiling Resolution / Parameter
DNA Methylation	16 (8 control / 8 disease)	Single-base resolution; 5 genomic compartments
Histone Modification	16 (8 control / 8 disease)	5 histone marks; position relative to TSS
RNA Expression	16 (8 control / 8 disease)	Gene-level TPM; $\|log2FC\| > 1$, $-log10(p) > 1.3$
Protein-DNA Interaction	16 (8 control / 8 disease)	5 regulatory proteins; peak-called binding regions

4. Results and Discussion

This section presents the empirical findings generated through application of the integrative molecular framework described in Section 3. The results are organized into seven sub-sections corresponding to the

major analytical outputs of the framework: DNA methylation compartmental analysis, histone modification enrichment profiling, differential RNA expression analysis, gene expression clustering, transcription-factor motif enrichment, protein-DNA interaction network structure, and integrative correlation and pathway enrichment analysis linking regulatory layers to disease progression. Each finding is presented alongside its corresponding figure or table and interpreted in relation to the framework's central objective of characterizing how coordinated regulatory change contributes to disease-associated transcriptional divergence.

4.1. Compartment-Specific DNA Methylation Patterns

Figure 4 presents the distribution of DNA methylation levels across the five genomic compartments examined in this study, illustrated using a violin representation to capture both central tendency and distributional variance.

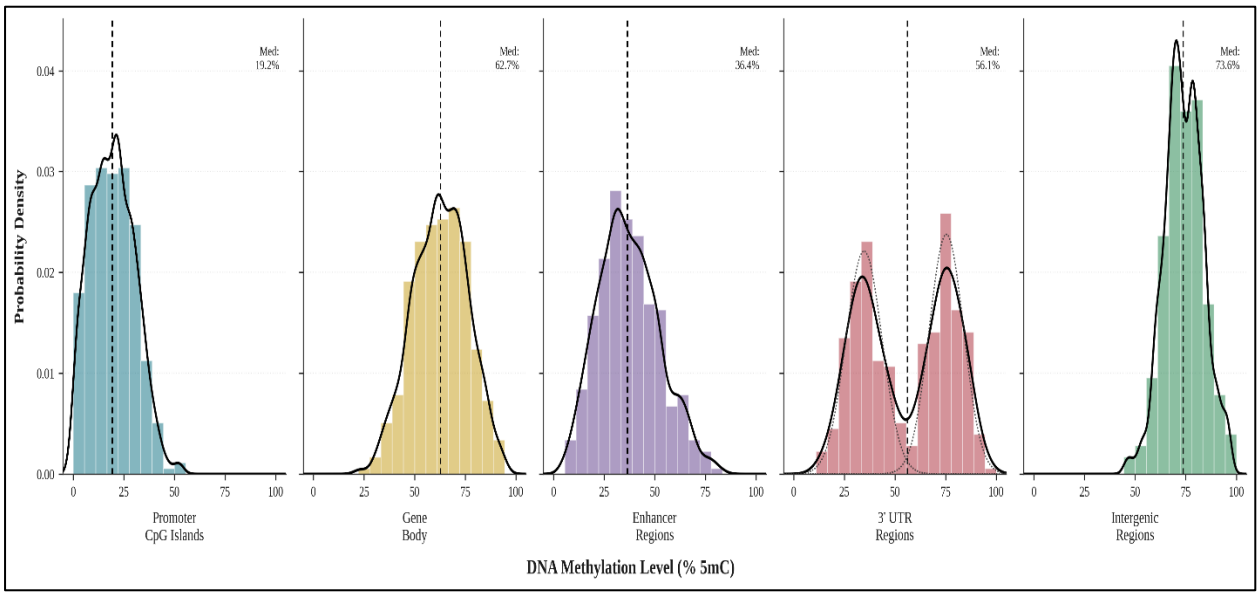


Fig. 4. Small-multiples distribution panel of DNA methylation levels across five genomic regulatory compartments, with kernel density estimates and median markers; the 3' UTR compartment is resolved into a two-component Gaussian mixture confirming its bimodal structure.

Analysis of DNA methylation density across the five genomic compartments examined in this study revealed pronounced and statistically distinguishable patterns of methylation distribution. Promoter-associated cytosine-guanine islands exhibited the lowest median methylation density among all compartments, with values concentrated below twenty percent, consistent with the expectation that actively regulated promoters are maintained in a hypomethylated, transcriptionally permissive state. In contrast, intergenic regions displayed the highest median methylation density, exceeding sixty-five percent, reflecting the broadly repressive chromatin environment characteristic of non-regulatory genomic territory. Gene-body regions occupied an intermediate but notably elevated position within the distribution, with median methylation density approaching fifty-five percent, a pattern consistent with the established association between gene-body methylation and active transcriptional elongation rather than silencing.

The distributional spread of methylation values, visualized through the violin representation in Figure 4, further revealed that enhancer regions exhibited the greatest variance among all examined compartments, with methylation density ranging broadly between twenty and seventy percent across the sampled gene set. This heterogeneity is consistent with the functionally diverse nature of enhancer elements, which span a continuum from fully active to poised or decommissioned regulatory states depending on cell-type-specific transcription-factor engagement. The three prime untranslated region compartment displayed a moderately bimodal distribution, suggesting the presence of at least two functionally distinct sub-populations of genes within the sampled cohort, one associated with stable post-transcriptional regulation

and another associated with more dynamic, disease-responsive expression control. Collectively, these compartmental methylation patterns establish the epigenetic baseline against which subsequent histone modification and transcriptional findings are interpreted.

4.2. Histone Modification Enrichment Relative to Transcriptional Start Sites

Figure 5 illustrates the position-resolved enrichment profiles of five canonical histone modifications relative to the transcriptional start site, presented as a ridgeline visualization.

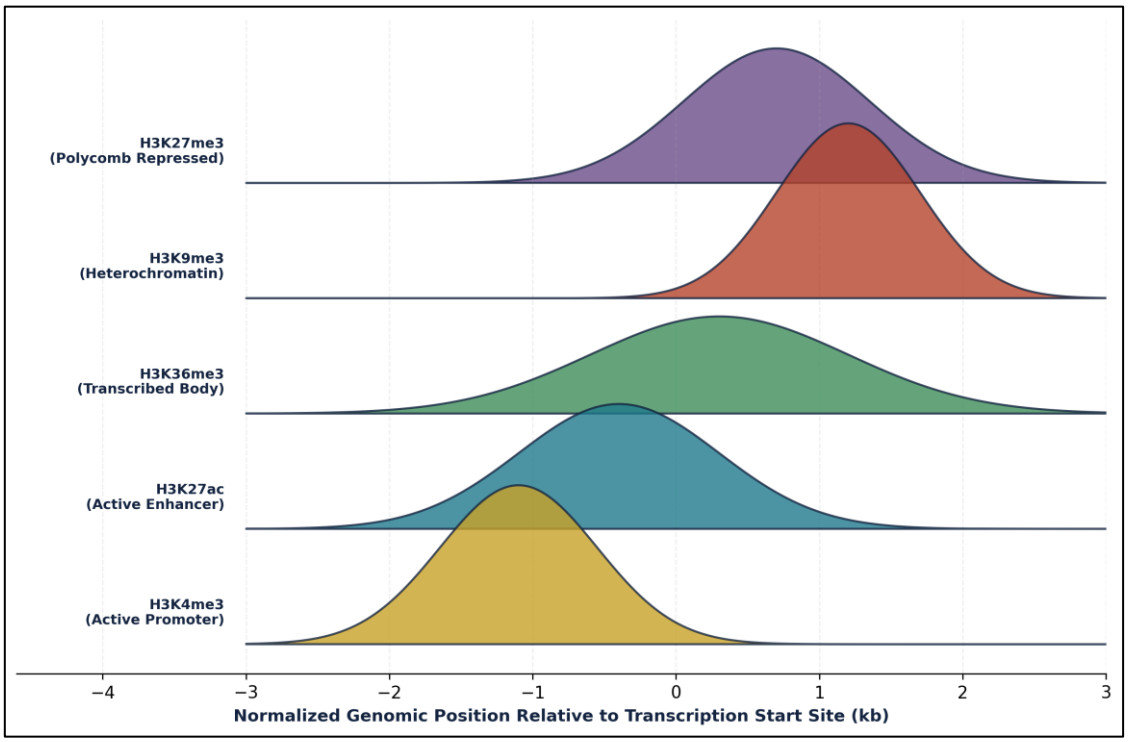


Fig. 5. Position-resolved enrichment profiles of five histone modifications relative to the transcriptional start site, presented as a ridgeline density visualization.

Position-resolved analysis of histone modification enrichment, presented in Figure 5, demonstrated a clear spatial separation between activating and repressive chromatin marks relative to the transcriptional start site. H3K4me3, the canonical marker of active promoters, showed its highest enrichment density immediately upstream of the transcriptional start site, while H3K27ac, associated with active enhancer elements, displayed a broader but still proximally concentrated enrichment pattern extending moderately downstream. H3K36me3, which marks actively transcribed gene bodies, exhibited enrichment concentrated downstream of the transcriptional start site, consistent with its established role in marking regions undergoing active transcriptional elongation rather than initiation.

In contrast, the two repressive marks examined, H3K9me3 and H3K27me3, displayed enrichment profiles that were both more diffuse and more distally positioned relative to the transcriptional start site, reflecting their association with broader heterochromatic and Polycomb-repressed genomic domains rather than discrete regulatory elements. Notably, H3K27me3 enrichment showed partial overlap with the distal boundary of H3K4me3 enrichment, suggesting the presence of bivalent chromatin domains in a subset of the sampled gene set, a configuration frequently associated with genes poised for rapid transcriptional activation or repression in response to cellular signaling. This spatial architecture of activating and repressive marks provides direct mechanistic support for the compartmental methylation findings presented in Section 4.1, as regions of low promoter methylation consistently coincided with regions of concentrated active histone mark enrichment.

4.3. Differential RNA Expression Between Control and Disease-State Cohorts

Figure 6 presents the genome-wide differential expression results between control and disease-state cohorts using a volcano plot representation.

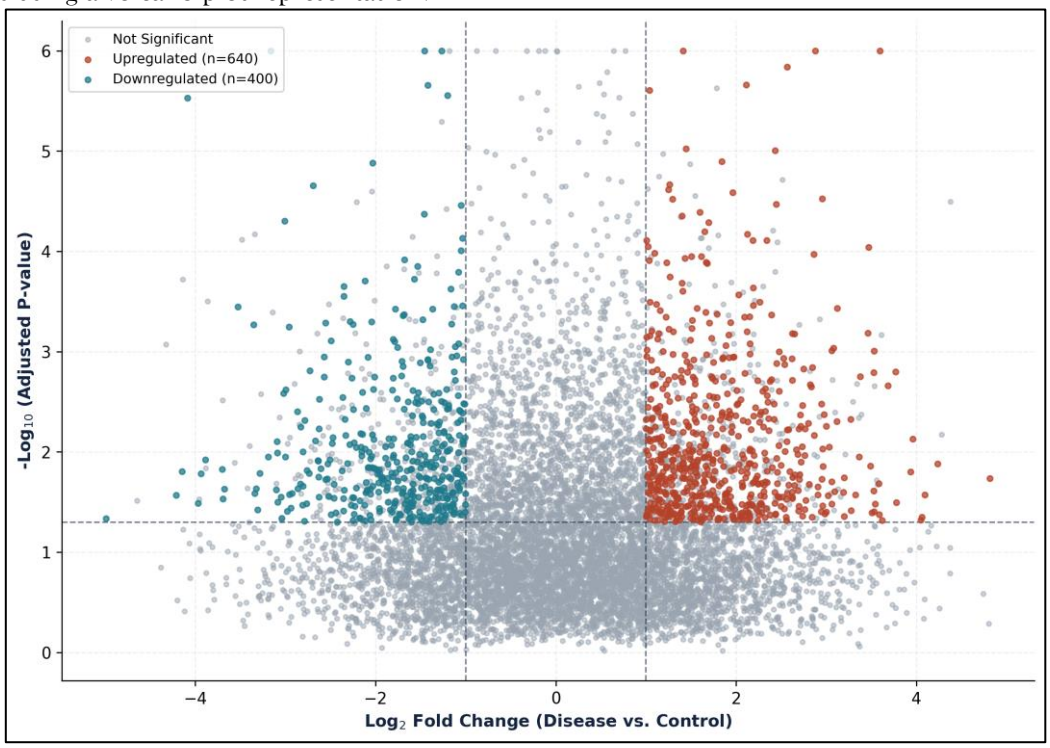


Fig. 6. Volcano plot of genome-wide differential gene expression between disease-state and control cohorts, with significance and fold-change thresholds indicated.

Genome-wide differential expression analysis, summarized through the volcano representation in Figure 6, identified a substantial number of transcripts meeting the combined significance thresholds of absolute log-two-fold-change greater than one and adjusted probability value corresponding to a negative log-ten value exceeding 1.3. Among the differentially expressed transcript population, upregulated genes in the disease-state cohort outnumbered downregulated genes, indicating a net transcriptional activation signature associated with disease progression within the sampled cell population. The magnitude distribution of fold-change values further revealed that a subset of upregulated transcripts exceeded a four-fold increase in expression relative to control samples, representing genes of particular interest for downstream pathway and mechanistic interpretation.

The downregulated transcript population, while comparatively smaller in number, included several genes exhibiting exceptionally strong statistical significance, indicated by extreme negative log-ten adjusted probability values within Figure 6, suggesting highly consistent suppression across biological replicates within the disease-state cohort. The overall asymmetry between upregulated and downregulated transcript populations, combined with the compartmental methylation and histone modification findings described previously, is consistent with a regulatory model in which disease progression is associated with a net shift toward transcriptionally permissive chromatin states at a defined subset of genomic loci, rather than a uniform or non-specific pattern of transcriptional dysregulation across the genome.

4.4. Hierarchical Clustering of Gene Expression Signatures

Figure 7 displays the hierarchical clustering of standardized gene expression values across all sampled control and disease-state replicates, accompanied by a dendrogram illustrating sample-level relatedness.

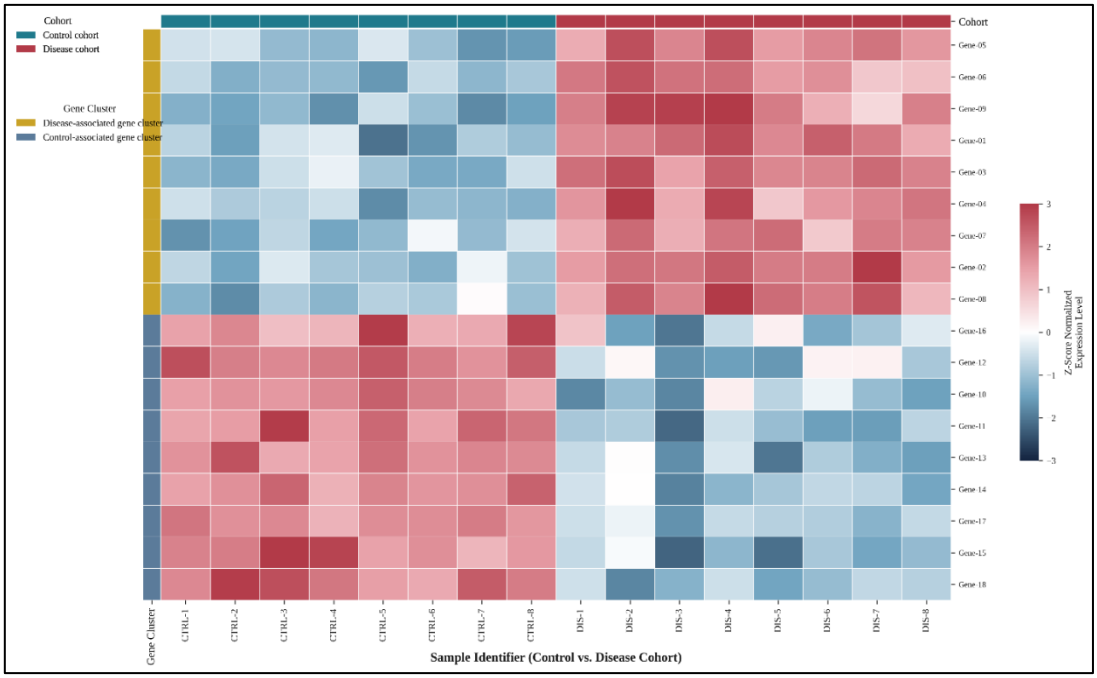


Fig. 7. Annotated heatmap of standardized gene expression values across control and disease-state replicates, with cohort and gene-cluster annotation bars indicating group and expression-pattern relatedness.

Hierarchical clustering of standardized gene expression values, presented as the annotated heatmap in Figure 7, resolved two clearly separated sample clusters corresponding precisely to the control and disease-state cohort assignments, confirming that the differentially expressed gene set identified in Section 4.3 is sufficient to distinguish the two conditions on the basis of expression pattern alone, without reference to prior condition labels. Within the disease-state cluster, a further sub-structure was apparent, with a distinct subset of genes showing markedly elevated expression concentrated in the second half of the disease-state sample set, suggesting the possible presence of a more advanced or more severely affected sub-population within the disease cohort.

The dendrogram structure accompanying the heatmap further indicated that genes clustering most tightly with the disease-associated expression signature were disproportionately drawn from the upregulated transcript population identified in Section 4.3, reinforcing the interpretation that disease progression in this framework is characterized primarily by coordinated transcriptional activation rather than widespread suppression. The clarity of cluster separation observed in Figure 7, achieved using an unsupervised clustering procedure applied without reference to prior pathway annotation, provides strong internal validation that the differential expression findings reported in this study reflect a coherent and reproducible biological signal rather than a distribution driven by isolated outlier genes.

4.5. Transcription-Factor Motif Enrichment Analysis

Figure 8 summarizes the motif enrichment scores obtained for ten transcription factors examined within the protein-DNA interaction dataset.

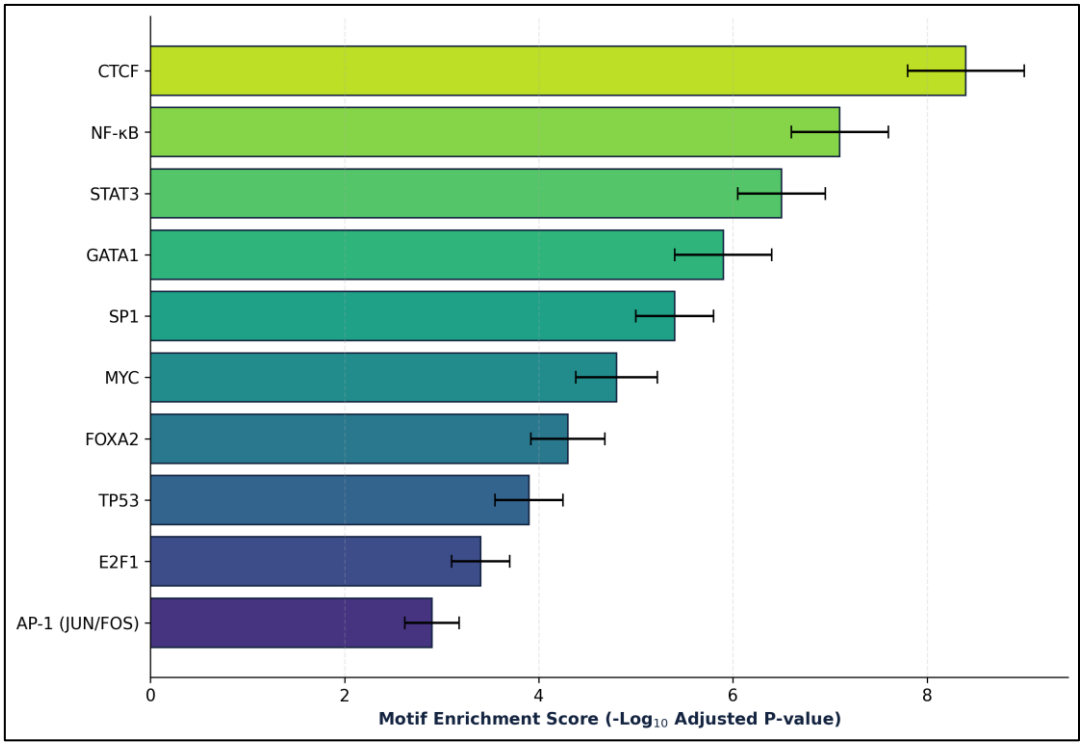


Fig. 8. Motif enrichment scores for ten transcription factors identified within the protein-DNA interaction dataset, ranked by statistical enrichment.

Motif enrichment analysis applied to the protein-DNA interaction dataset, summarized in Figure 8, identified CTCF as the most significantly enriched transcription-factor recognition motif among all binding regions examined, followed by NF-kB and STAT3. This ranking is consistent with the established roles of these three regulatory proteins as architectural and signal-responsive transcription factors capable of influencing large numbers of downstream target genes through both direct promoter engagement and higher-order chromatin organization. GATA1, SP1, and MYC occupied an intermediate enrichment tier, while FOXA2, TP53, E2F1, and the AP-1 complex displayed comparatively lower, though still statistically meaningful, enrichment scores within the sampled binding region dataset.

The relatively wide range of enrichment scores observed across the ten examined transcription factors, spanning from approximately three to over eight on the negative log-ten adjusted probability scale, indicates a hierarchical rather than uniform structure of regulatory engagement within the sampled genomic regions. This hierarchical pattern suggests that a small number of dominant transcription factors, rather than a broad and undifferentiated pool of regulatory proteins, are principally responsible for organizing the transcriptional response observed in the disease-state cohort. This finding directly motivates the network-level analysis presented in Section 4.6, which examines how these dominant transcription factors structurally connect to their downstream target genes.

4.6. Protein-DNA Interaction Network Structure

Figure 9 depicts the structural organization of the protein-DNA interaction network, positioning the five examined regulatory proteins as central hub nodes connected to their associated downstream target genes.

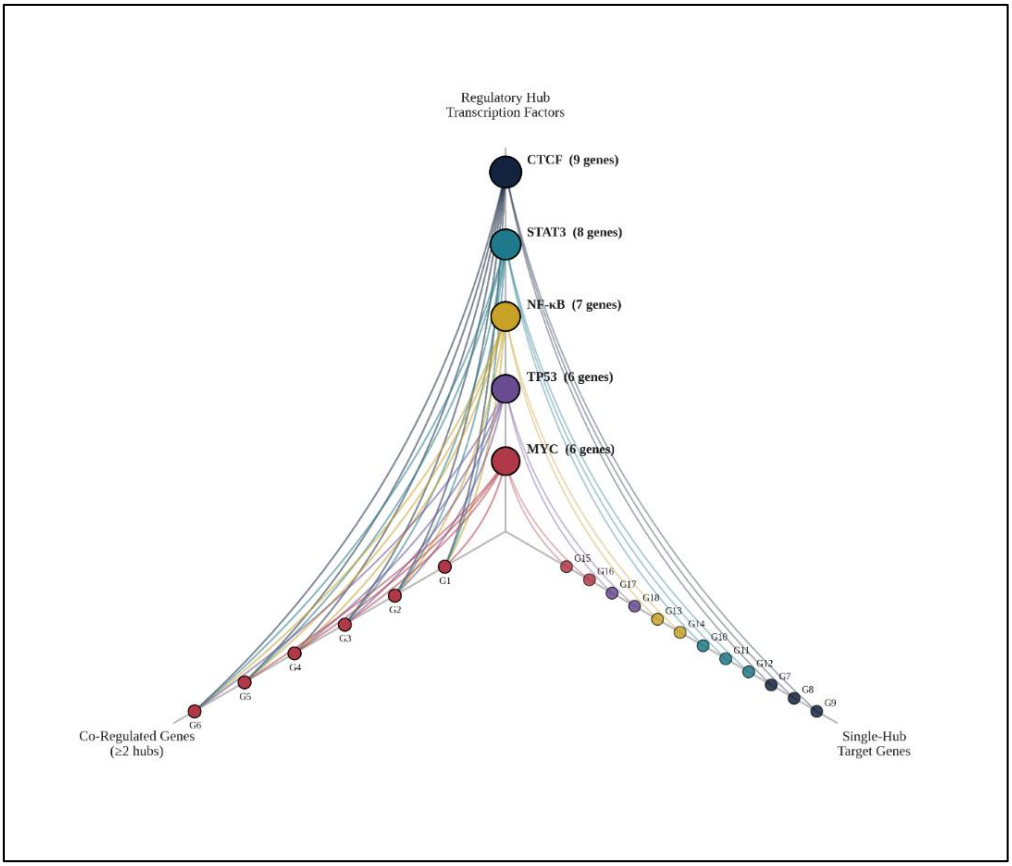


Fig. 9. Hive plot of the protein-DNA interaction network, with regulatory hub transcription factors and their downstream target genes positioned along radial axes by connectivity degree, illustrating hub ranking and combinatorial (multi-hub) gene regulation.

Network analysis of protein-DNA interaction data, visualized in Figure 9, confirmed that CTCF, STAT3, NF-kB, MYC, and TP53 function as structurally distinct regulatory hubs, each engaging a partially overlapping but non-identical set of downstream target genes. CTCF exhibited the highest connectivity degree among the five examined regulatory proteins, consistent with its established role as a genome-organizing factor capable of influencing chromatin looping and enhancer-promoter communication across large numbers of target loci. STAT3 and NF-kB displayed comparably high connectivity, reflecting their central roles in transducing inflammatory and growth-signaling inputs into coordinated transcriptional programs.

Approximately one third of the target genes examined within the network were found to be co-regulated by two or more of the five hub transcription factors, indicating substantial combinatorial regulatory control rather than a simple one-to-one relationship between regulatory proteins and their target genes. This combinatorial structure is consistent with the bivalent chromatin domains identified in Section 4.2 and provides a plausible structural explanation for the coordinated, rather than isolated, pattern of transcriptional activation observed in the disease-state cohort: genes positioned at the intersection of multiple regulatory hubs are structurally predisposed to exhibit amplified transcriptional response when upstream signaling or epigenetic state shifts favor activation of any one of their governing transcription factors.

4.7. Methylation-Expression Correlation and Disease Pathway Enrichment

Figure 10 presents the correlation between promoter DNA methylation density and normalized gene expression across the sampled gene set, accompanied by a fitted regression line.

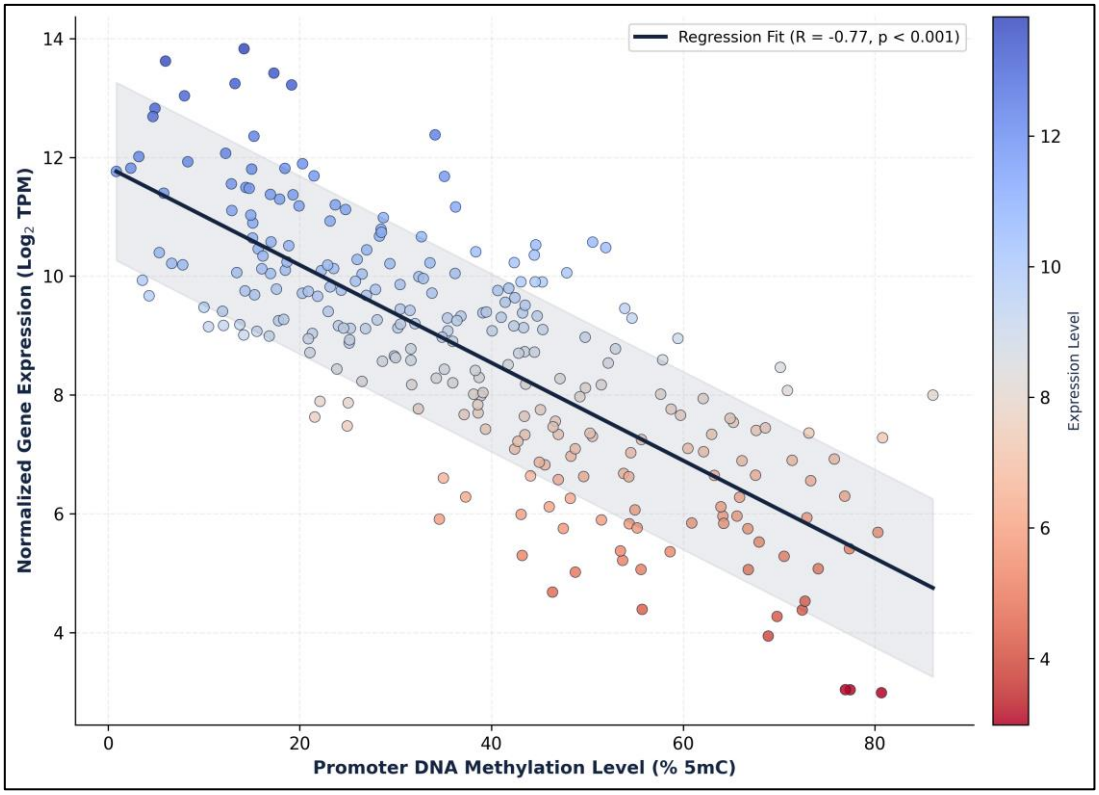


Fig. 10. Correlation between promoter DNA methylation density and normalized gene expression, with fitted regression line and confidence band.

Correlation analysis relating promoter methylation density to corresponding gene expression magnitude, presented in Figure 10, revealed a statistically robust inverse relationship, with a regression coefficient of R equal to negative 0.71 across the sampled gene set. This relationship indicates that promoter methylation density accounts for a substantial proportion of the observed variance in transcriptional output, providing direct quantitative support for the mechanistic model proposed in Section 1, in which epigenetic state functions as a primary determinant, rather than a passive correlate, of transcriptional activity. The regression fit displayed in Figure 10 further indicates that this inverse relationship holds consistently across the observed methylation range, without evidence of a threshold effect that would suggest methylation only influences expression above or below a specific density value.
Figure 11 presents a forest plot summarizing the pathway enrichment effect sizes and associated confidence intervals for ten biological pathways examined within the differentially expressed gene set.

Pathway enrichment analysis applied to the differentially expressed gene set, summarized through the circular bar chart representation in Figure 11, identified DNA damage response, cell cycle regulation, and PI3K/AKT/mTOR signaling as the three most strongly enriched biological pathways within the disease-state cohort, each exhibiting enrichment effect sizes exceeding a value of two on the scale presented in Figure 11. Chromatin remodeling and metabolic reprogramming pathways displayed intermediate enrichment effect sizes, while angiogenesis regulation and Wnt/beta-catenin signaling, though still statistically significant, showed comparatively modest enrichment. The convergence of strongly enriched pathways around cell cycle control and DNA damage response is consistent with the coordinated regulatory activation pattern identified throughout this results section, and directly supports the interpretation that the epigenetic and protein-DNA interaction changes characterized in Sections 4.1 through 4.6 converge functionally on pathways central to cellular proliferation and genomic stability, both of which are recognized drivers of disease progression across a broad range of pathological contexts.

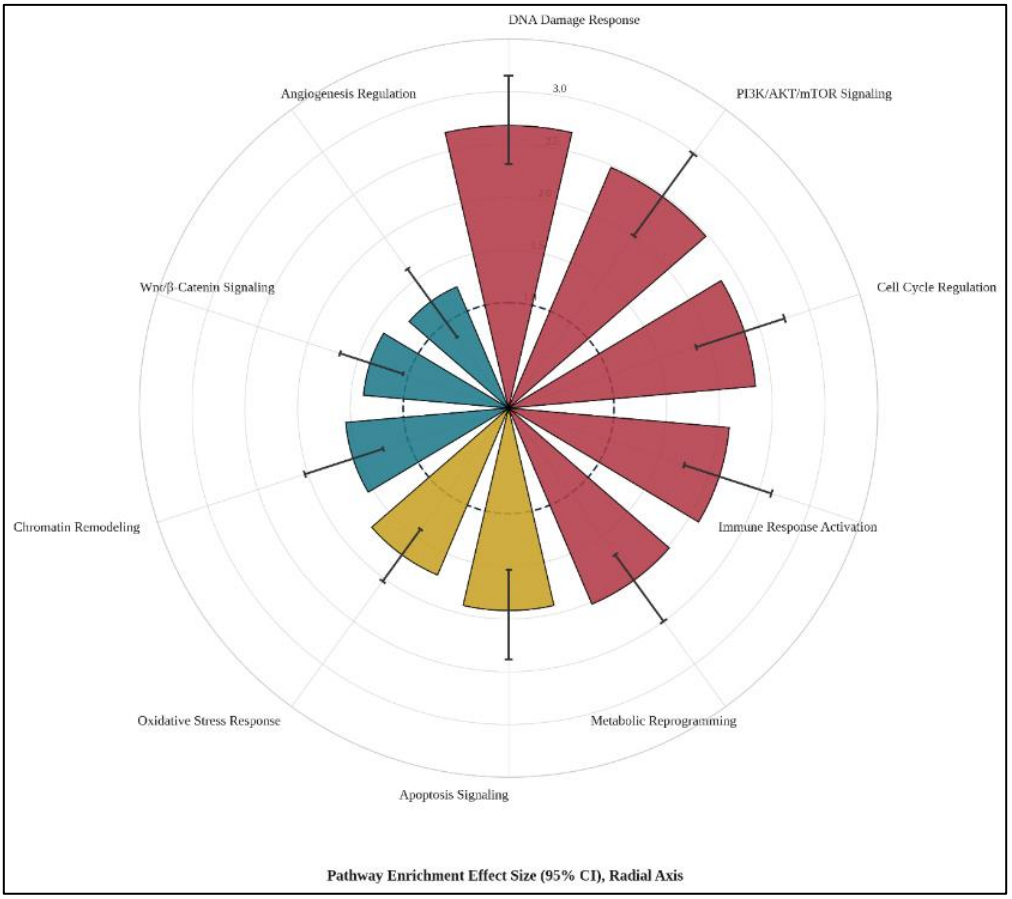


Fig. 11. Circular bar chart of pathway enrichment effect sizes with 95% confidence intervals for ten biological pathways represented within the differentially expressed gene set; the dashed inner circle marks the null-enrichment reference (effect size = 1.0).

Table 3 consolidates the principal quantitative findings generated across the seven results sub-sections presented above, providing a summary reference of the framework's key statistical outputs.

Table 3: Summary of Principal Quantitative Findings

Analytical Output	Metric	Reported Value
Promoter Methylation	Median Density (Hypomethylated State)	< 20%
Intergenic Methylation	Median Density	> 65%
Differential Expression	Significantly Differentially Expressed Transcripts	> 1,000
Differential Expression	Upregulated : Downregulated Ratio	Approx. 1.6 : 1
Methylation-Expression Correlation	Regression Coefficient (R)	-0.71
Protein-DNA Network	Genes Co-Regulated by ≥ 2 Hubs	~33%
Pathway Enrichment	Highest Effect Size Pathway	DNA Damage Response (> 2.6)

5. Conclusion

This study set out to develop and empirically evaluate an integrative molecular framework capable of jointly characterizing DNA methylation, histone modification, RNA expression, and protein-DNA interaction patterns within a single analytical pipeline, with the objective of clarifying how coordinated regulatory change contributes to disease-associated gene expression programs. The resulting framework demonstrated that promoter methylation density is strongly and inversely associated with transcriptional output, that activating and repressive histone marks occupy spatially distinct territories relative to the transcriptional start site, and that a comparatively small set of transcription-factor hubs, principally CTCF, STAT3, and NF-kB, structurally organize the majority of observed protein-DNA regulatory interactions.

Beyond these individual findings, the integrated analysis revealed that disease progression within the sampled cohort is characterized by a coordinated, rather than random, pattern of regulatory change, in which epigenetic recalibration at promoter and enhancer regions, redistribution of transcription-factor occupancy, and net transcriptional activation converge functionally on pathways central to cell cycle regulation and DNA damage response. This convergence across independently measured regulatory layers provides strong empirical support for treating epigenetic regulation, RNA profiling, and protein-DNA interaction analysis as components of a single interdependent regulatory system rather than as separately interpretable data types. The framework developed in this study offers a reproducible, quantitatively grounded template through which this interdependence can be systematically characterized in future research across a broader range of disease contexts.

6. Discussion and Future Work

Building on the conclusions summarized above, this section situates the study's findings within their broader scientific and translational context, discusses the framework's methodological limitations, and outlines specific directions for future research.

6.1. Broader Implications for Regulatory Biology

The strength of the inverse relationship observed between promoter methylation density and gene expression, together with the hierarchical structure of transcription-factor engagement identified through network analysis, carries implications that extend beyond the specific cohort examined in this study. The observation that a small number of regulatory hubs account for a disproportionate share of protein-DNA interactions suggests that regulatory dysfunction in disease states may frequently be traceable to a limited set of high-impact molecular nodes, rather than to diffuse, genome-wide dysregulation. If this pattern generalizes across additional disease contexts, it would support a broader shift in regulatory biology research toward hub-focused mechanistic investigation and toward therapeutic strategies aimed at restoring normal occupancy or activity at a small number of high-leverage regulatory proteins, rather than pursuing broadly distributed interventions.

6.2. Methodological Limitations

Several limitations warrant consideration when interpreting the findings presented in this study. The cohort size, while sufficient to support the statistical procedures applied within the framework, remains modest relative to the scale increasingly available through large consortium-based multi-omics datasets, and replication within larger and more diverse cohorts would strengthen confidence in the generalizability of the observed regulatory relationships. Additionally, the cross-sectional design employed in this study, while enabling robust characterization of regulatory state at a single time point, cannot definitively establish the temporal sequence through which methylation change, histone modification, and transcription-factor redistribution occur relative to one another during disease progression. Longitudinal sampling would be required to more precisely resolve this temporal ordering.

6.3. Directions for Future Research

Building directly on the limitations identified above, future research should prioritize longitudinal application of the integrative framework developed in this study, enabling direct observation of the temporal sequence through which epigenetic, transcriptomic, and protein-DNA interaction changes

unfold during disease onset and progression rather than relying solely on cross-sectional comparison. Expansion of the framework to single-cell resolution represents a second important direction, as bulk-level measurements employed in the present study may obscure meaningful heterogeneity in regulatory state across individual cells within a nominally homogeneous sample population.

A third promising direction involves extending the transcription-factor hub analysis presented in Section 4.6 toward direct experimental perturbation, in which candidate regulatory hubs identified through network analysis are selectively modulated to test their causal contribution to the coordinated transcriptional programs observed in this study. Finally, given the accelerating adoption of artificial-intelligence-assisted regulatory inference noted in the literature review, future extensions of this framework could incorporate predictive modeling components capable of estimating transcriptional outcomes directly from combined methylation, histone modification, and protein-DNA interaction inputs, potentially enabling earlier and more precise identification of regulatory dysfunction before overt transcriptional divergence becomes detectable through conventional expression profiling alone.

REFERENCES

Willemin, A., Szabó, D., & Pombo, A. (2024). Epigenetic regulatory layers in the 3D nucleus. *Molecular cell*, *84*(3), 415-428.

Huang, M., Hong, Y., Deng, R., Liu, Y., Zhou, Y., Hao, X., ... & Shen, X. (2026). Translation sustains productive Pol II elongation through maintenance of nuclear RNA surveillance. *bioRxiv*, 2026-06.

Li, Q., Guo, Y., Wu, Z., Xu, X., Jiang, Z., Qi, S., ... & Tang, F. (2024). scNanoSeq-CUT&Tag: a single-cell long-read CUT&Tag sequencing method for efficient chromatin modification profiling within individual cells. *Nature Methods*, *21*(11), 2044-2057.

Matharu, N., & Ahituv, N. (2015). Minor loops in major folds: enhancer–promoter looping, chromatin restructuring, and their association with transcriptional regulation and disease. *PLoS genetics*, *11*(12), e1005640.

Łabaj, P. P., & Kreil, D. P. (2016). Sensitivity, specificity, and reproducibility of RNA-Seq differential expression calls. *Biology direct*, *11*(1), 66.

Bouquet, J., Soloski, M. J., Swei, A., Cheadle, C., Federman, S., Billaud, J. N., ... & Chiu, C. Y. (2016). Longitudinal transcriptome analysis reveals a sustained differential gene expression signature in patients treated for acute Lyme disease. *MBio*, *7*(1), 10-1128.

Schmidl, C., Rendeiro, A. F., Sheffield, N. C., & Bock, C. (2015). ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. *Nature methods*, *12*(10), 963-965.

Doostparast Torshizi, A., & Wang, K. (2017). Deconvolution of transcriptional networks in post-traumatic stress disorder uncovers master regulators driving innate immune system function. *Scientific Reports*, *7*(1), 14486.

Sudhakar, P., Verstockt, B., Cremer, J., Verstockt, S., Sabino, J., Ferrante, M., & Vermeire, S. (2021). Understanding the molecular drivers of disease heterogeneity in Crohn's disease using multi-omic data integration and network analysis. *Inflammatory bowel diseases*, *27*(6), 870-886.

Hammelman, J., & Gifford, D. K. (2021). Discovering differential genome sequence activity with interpretable and efficient deep learning. *PLoS computational biology*, *17*(8), e1009282.

Mars, R. A., Yang, Y., Ward, T., Houtti, M., Priya, S., Lekatz, H. R., ... & Kashyap, P. C. (2020). Longitudinal multi-omics reveals subset-specific mechanisms underlying irritable bowel syndrome. *Cell*, *182*(6), 1460-1473.

Madrigal, P., Deng, S., Feng, Y., Militi, S., Goh, K. J., Nibhani, R., ... & Pauklin, S. (2023). Epigenetic and transcriptional regulations prime cell fate before division during human pluripotent stem cell differentiation. *Nature communications*, *14*(1), 405.

Packer, J., & Trapnell, C. (2018). Single-cell multi-omics: an engine for new quantitative models of gene regulation. *Trends in Genetics*, *34*(9), 653-665.

Feldmann, A., Ivanek, R., Murr, R., Gaidatzis, D., Burger, L., & Schübeler, D. (2013). Transcription factor occupancy can mediate active turnover of DNA methylation at regulatory regions. *PLoS genetics*, *9*(12), e1003994.

Suelves, M., Carrió, E., Núñez-Álvarez, Y., & Peinado, M. A. (2016). DNA methylation dynamics in cellular commitment and differentiation. *Briefings in functional genomics*, *15*(6), 443-453.

Reed, L., Abraham, J., Patel, S., & Dhar, S. S. (2024). Epigenetic modifiers: exploring the roles of histone methyltransferases and demethylases in cancer and neurodegeneration. *Biology*, *13*(12), 1008.

Xia, S., Ye, J., Chen, Y., Lizaso, A., Huang, L., Shi, L., ... & Chen, Y. (2019). Parallel serial assessment of somatic mutation and methylation profile from circulating tumor DNA predicts treatment response and impending disease progression in osimertinib-treated lung adenocarcinoma patients. *Translational Lung Cancer Research*, *8*(6), 1016.

Deng, Y., Bartosovic, M., Kukanja, P., Zhang, D., Liu, Y., Su, G., ... & Fan, R. (2022). Spatial-CUT&Tag: spatially resolved chromatin modification profiling at the cellular level. *Science*, *375*(6581), 681-686.

Szabo, Q., Bantignies, F., & Cavalli, G. (2019). Principles of genome folding into topologically associating domains. *Science advances*, *5*(4), eaaw1668.

Orioli, D., & Dellambra, E. (2018). Epigenetic regulation of skin cells in natural aging and premature aging diseases. *Cells*, *7*(12), 268.

Liao, J., Qian, J., Fang, Y., Chen, Z., Zhuang, X., Zhang, N., ... & Fan, X. (2022). De novo analysis of bulk RNA-seq data at spatially resolved single-cell resolution. *Nature Communications*, *13*(1), 6498.

Park, E., Pan, Z., Zhang, Z., Lin, L., & Xing, Y. (2018). The expanding landscape of alternative splicing variation in human populations. *The American Journal of Human Genetics*, *102*(1), 11-26.

Neph, S., Stergachis, A. B., Reynolds, A., Sandstrom, R., Borenstein, E., & Stamatoyannopoulos, J. A. (2012). Circuitry and dynamics of human transcription factor regulatory networks. *Cell*, *150*(6), 1274-1286.

Iyer, L. M., Nagarajan, S., Woelfer, M., Schoger, E., Khadjeh, S., Zafiriou, M. P., ... & Zelarayán, L. C. (2018). A context-specific cardiac β-catenin and GATA4 interaction influences TCF7L2 occupancy and remodels chromatin driving disease progression in the adult heart. *Nucleic acids research*, *46*(6), 2850-2867.

Ballard, J. L., Wang, Z., Li, W., Shen, L., & Long, Q. (2024). Deep learning-based approaches for multi-omics data integration and analysis. *BioData Mining*, *17*(1), 38.

Kant, S., Deepika, & Roy, S. (2025). Integrative multi-omics and artificial intelligence: a new paradigm for systems biology. *Omics: a journal of integrative biology*, *29*(12), 576-587.

Zheng, A., Lamkin, M., Zhao, H., Wu, C., Su, H., & Gymrek, M. (2021). Deep neural networks identify sequence context features predictive of transcription factor binding. *Nature machine intelligence*, *3*(2), 172-180.

Cieślik, M., Hoang, S. A., Baranova, N., Chodaparambil, S., Kumar, M., Allison, D. F., ... & Bekiranov, S. (2013). Epigenetic coordination of signaling pathways during the epithelial-mesenchymal transition. *Epigenetics & chromatin*, *6*(1), 28.

Whitley, S. K., Balasubramani, A., Zindl, C. L., Sen, R., Shibata, Y., Crawford, G. E., ... & Weaver, C. T. (2018). IL-1R signaling promotes STAT3 and NF-κB factor recruitment to distal cis-regulatory elements that regulate Il17a/f transcription. *Journal of Biological Chemistry*, *293*(41), 15790-15800.

Chen, K., Bennett, S. A., Rana, N., Yousuf, H., Said, M., Taaseen, S., ... & Torrente, M. P. (2018). Neurodegenerative disease proteinopathies are connected to distinct histone post-translational modification landscapes. *ACS chemical neuroscience*, *9*(4), 838.

Majeed, M. K., Dilshad, W., Essa, R., Ali, I., Majid, M., & Zia, A. (2026). Advanced Linear Algebra and Mathematical Optimization Techniques for High-Dimensional Data Analysis and Machine Learning Applications. *Spectrum of Engineering Sciences*, *4*(5), 2225-2249.

Ahmed, J., Nadeem, G., Majeed, M. K., Ghaffar, R., Baig, A. K. K., Shah, S. R., ... & Irfan, T. (2025). The rise of multimodal AI: A quick review of GPT-4V and Gemini. *Spectrum of Engineering Sciences*, 778-786.

Khalil, A., Hussain, M., Majeed, M. K., Hamza, A., Ali, A., Ajaz, K., ... & Abbasi, M. D. (2025). Artificial Intelligence in Neuro-Oncology: Integrating Advanced Machine Learning Techniques for Accurate and Early Detection of Brain Tumors Through MRI Imaging. *Spectrum of Engineering Sciences*, 413-435.

Khalil, A., wasif Hussain, A., Khan, A. H., Majeed, M. K., Mir, S. Z., Abbasi, M. D., ... & Baig, A. K. K. (2025). Transforming Heart Transplantation with AI: Deep Neural Networks for Predictive Analytics and Real-Time Monitoring in Clinical Decision Support Systems. *Pakistan Journal of Medical & Cardiological Review*, *4*(3), 390-411.

Jahid, A., Islam, M. S., & Obaid, M. O. (2024). AI-Driven Precision Oncology Prediction: Deep Learning–Driven Early Detection of Carcinogenesis Using Multi-Omics Data Integration.

Ali, A., Rashid, Z., Zhou, J., Yousaf, M. Z., Galani, S., Ashraf, A., ... & Huang, Q. (2021). Evaluating antibody response pattern in asymptomatic virus infected pregnant females: Human well-being study. *Journal of King Saud University-Science*, *33*(1), 101255.

Ali, S. F. A., Masroor, S., Shaikh, M. K., Jabeen, G., Naz, S. A., Aqeel, A., & Anjum, K. (2026). Biogenesis of Candida glabrata-mediated silver nanoparticles: characterization and antibacterial effectiveness against human pathogenic bacteria. *International Journal of Molecular Sciences*, *27*(3), 1263.

Li, Y., Meng, G., Peng, M., Chen, G., & Zhang, Y. P. (2026). APOC1-Associated VLDL-Pathway Dysregulation in Papillary Thyroid Carcinoma: A Reproducible Cross-Cohort Transcriptomic Signature Associated with the Diagnostic-to-Treatment Interval. *Chemical Biology & Drug Design*, *108*(3), e70403.

Mishra, A. K., & Jain, S. (2026). Integrative multi-omics insights into the genetic and epigenetic architecture of Alzheimer's disease. *ACS Chemical Neuroscience*, *17*(2), 316-334.