Deep Learning Analysis of Skeletal Muscle Gene Expression and Splicing in Exercise-Induced Metabolic Adaptation

Hanfei Wei
Department of Computer Science, Colorado State University, Fort Collins, CO, USA.
hanfeiwei91@colostate.edu

Abstract

Skeletal muscle is a highly plastic tissue that orchestrates systemic metabolic adaptation to exercise through coordinated changes in gene expression and alternative splicing. The complexity of these regulatory layers has historically outstripped the analytical capacity of conventional bioinformatics, but the advent of deep learning architectures is transforming the study of exercise-induced transcriptomic remodeling. This paper presents a system-level analysis of how deep neural networks can be leveraged to model the interplay between gene expression and splicing in human skeletal muscle during metabolic adaptation. We examine the architectural choices—ranging from convolutional and recurrent networks to attention-based transformers—that are particularly suited for capturing the sequence grammar of splicing and the dynamic patterns of expression. Beyond algorithm design, the paper addresses the infrastructure, data governance, and scalability challenges that arise when building large-scale, multi-omic data commons for exercise genomics. We discuss how federated learning, differential privacy, and fairness-aware model design can mitigate the risks of population bias and inequitable precision health outcomes. The integration of proton-sensing molecular reporters, transcriptomic deep learning, and metabolic phenotyping is explored as a frontier for capturing the ionic-stress dimension of exercise physiology. The analysis further evaluates the translational implications of these models for personalized exercise prescription and the policy frameworks required to ensure equitable access, algorithmic transparency, and long-term sustainability of AI-driven metabolic health infrastructure. Throughout, we emphasize structural trade-offs between model complexity and interpretability, the robustness of splicing predictions across ancestrally diverse cohorts, and the ethical responsibilities that accompany the deployment of deep learning in health-related behavioral interventions.

Keywords

deep learning, skeletal muscle, gene expression, alternative splicing, exercise metabolism, systems biology, data governance, fairness.

1. Introduction

Skeletal muscle constitutes approximately forty percent of human body mass and serves as the primary site for postprandial glucose disposal and fatty acid oxidation. In response to acute and chronic exercise, muscle fibers undergo profound transcriptional reprogramming that enhances oxidative capacity, mitochondrial biogenesis, and substrate flexibility, collectively termed metabolic adaptation. For decades, transcriptomic studies interrogated these adaptations through microarrays and bulk RNA sequencing, yielding catalogues of differentially expressed genes and enriched pathways. However, it has become increasingly

clear that gene expression levels alone cannot explain the full scope of metabolic plasticity; alternative splicing generates multiple mRNA isoforms from a single gene, often with divergent or even antagonistic metabolic functions. Understanding how exercise reshapes the splicing landscape and how genetic variation modulates both expression and splicing remains a formidable challenge that demands computational models capable of deciphering non-linear sequence-function relationships. Conventional statistical frameworks for differential splicing, such as those relying on linear mixed models or exon-centric counting, struggle to integrate the combinatorial and context-dependent nature of splice site selection, especially when confronted with the vast potential interaction space between single nucleotide polymorphisms, epigenetic marks, and RNA-binding protein gradients.

Deep learning has recently emerged as a transformative paradigm for biological sequence analysis, offering the capacity to learn hierarchical representations directly from raw genomic or transcriptomic data without explicit feature engineering. Methods such as SpliceAI have demonstrated that deep convolutional neural networks can predict splice junctions from pre-mRNA sequence with near-experimental accuracy, while graph-based and transformer architectures have enabled tissue-specific splicing code inference. In parallel, sequence-to-activity models like Enformer and DeepSEA have shown that long-range regulatory elements can be integrated into gene expression prediction. Yet, the application of these models to exercise-induced skeletal muscle adaptation remains nascent, partly because the field has lacked large, densely phenotyped human muscle multi-omic datasets that couple acute exercise challenges, longitudinal training interventions, and high-resolution splicing measurements. Recent studies, including comprehensive profiling of transcriptomic and splice variant responses to exercise and diet-induced weight loss, are beginning to fill this gap by linking single-nucleotide polymorphisms to variability in both transcript abundance and isoform ratios [4]. This paper takes a systems perspective, arguing that realizing the full potential of deep learning for muscle metabolic adaptation requires an integrative socio-technical framework that addresses not only model architecture but also data infrastructure, fairness, robustness, and governance.

The following sections unpack these dimensions from a structural and infrastructural vantage point. We explore how neural network architecture choices are shaped by the intrinsic properties of splicing and expression data, such as the exponential isoform complexity, the long-range dependencies imposed by chromatin organization, and the sample-limited nature of human exercise intervention trials. We then move to the system-level challenges of building federated, privacy-preserving data ecosystems that can sustain iterative model improvement across ancestrally diverse populations. The discussion extends to the robustness of deep splicing predictors when deployed on cohorts with different genetic backgrounds and environmental exposures, and we examine fairness implications for AI-guided exercise recommendations. Finally, we reflect on the convergence of transcriptomic deep learning with novel molecular reporters, including genetically encoded ionic-stress sensors, which may open a window into the proton-driven signaling pathways that link muscle activity, metabolic stress, and systemic sleep regulation [15].

2. Architectural Trade-Offs in Deep Learning Models for Skeletal Muscle Transcriptomes

The choice of deep learning architecture for modeling both gene expression and alternative splicing is dictated by the distinct topological and sequential characteristics of these molecular processes. Alternative splicing is fundamentally a problem of sequence grammar: the

recognition of splice sites, branch points, and auxiliary cis-regulatory motifs dispersed across intronic and exonic regions, where the relative positioning and combinatorial interaction of motifs determines isoform outcome. Convolutional neural networks (CNNs) have proven remarkably effective at capturing local sequence motifs through sliding filters, and when dilated convolutions are used, they can extend their receptive field to hundreds or even thousands of nucleotides without quadratic parameter growth. This makes dilated CNNs a computationally efficient backbone for predicting splicing across entire gene loci, as demonstrated by the SpliceAI framework. However, splicing decisions are also influenced by long-range chromatin features, transcriptional kinetics, and RNA secondary structure, which are not well captured by purely local convolution operations. Recurrent neural networks, such as long short-term memory units, can in principle model sequential dependencies across an entire transcript, but they suffer from vanishing gradients and practical difficulties in scaling to ultra-long muscle genes such as TTN or NEB, which span over a hundred kilobases.

Transformer architectures, with their self-attention mechanism, offer a compelling alternative by computing pairwise interactions between all positions in a sequence, enabling the model to directly attend to distant regulatory elements irrespective of genomic distance. The Enformer model, originally designed for gene expression prediction from DNA sequence, uses a combination of convolutional downsampling and transformer layers to integrate regulatory information from up to two hundred kilobases. Adapting such architectures to the joint prediction of expression and splicing in response to exercise presents specific trade-offs. On one hand, the attention mechanism can capture the coordinated regulation of splicing factors and transcription factors that are themselves transcriptionally induced by exercise, such as PGC-1α and MEF2. On the other hand, the quadratic complexity of full self-attention becomes prohibitive when the model must process multi-kilobase pre-mRNA sequences, especially if the goal is to perform tissue-specific and condition-specific predictions across hundreds of muscle samples. Recent hybrid designs that combine dilated convolutions with sparse attention or state-space models are beginning to bridge this gap, offering a linear scaling path without sacrificing long-range sensitivity.

An additional architectural consideration is the multi-task learning paradigm. Exercise-induced metabolic adaptation simultaneously alters the expression of hundreds of genes and the splicing of thousands of isoforms, often in a coordinated manner. Training separate models for expression and splicing misses the opportunity to learn shared latent representations that reflect the underlying cellular state. A system that jointly predicts transcript abundance and isoform usage from a combination of DNA sequence, genotype, and epigenetic context can exploit the fact that the same regulatory variants may subtly shift both polymerase II elongation rate and splicing factor recruitment. Architecturally, this can be implemented via a shared encoder backbone that feeds into task-specific decoding heads, with carefully weighted loss functions balancing the scales of expression and splicing outputs. The risk is that the splicing head, which typically operates on a much higher-dimensional output space, may dominate gradient updates and destabilize the expression predictions. Gradient surgery techniques and uncertainty-weighted losses offer means to navigate this trade-off while preserving the ability to detect cross-layer regulatory interactions.

3. Integrative Modeling of Gene Expression and Alternative Splicing in Metabolic Adaptation

The integration of gene expression and splicing into a unified deep learning framework goes beyond architectural convenience; it addresses a biological reality where protein isoform

diversification is a cornerstone of metabolic flexibility. During endurance exercise, for instance, the transcriptional co-activator PGC-1α undergoes alternative promoter usage and splicing that produce isoforms with distinct tissue distributions and metabolic functions, including a truncated variant that retains the ability to induce mitochondrial genes while evading negative feedback. Similarly, the AMP-activated protein kinase signaling cascade, which is central to exercise-induced energy sensing, regulates a suite of splicing factors such as RBM20 and SRSF3 that drive isoform switches in metabolic enzymes like pyruvate kinase and carnitine palmitoyltransferase. A deep learning model that ingests post-exercise muscle RNA-seq data and simultaneously predicts expression levels and splice junction usage can learn to associate specific isoform transitions with improvements in insulin sensitivity or maximal oxygen uptake. This requires not only accurate molecular predictions but also careful phenotypic anchoring, where the system is trained on paired transcriptomic and clinical phenotyping data from randomized exercise trials.

Constructing such integrative models raises substantial data representation challenges. The output space for splicing is inherently combinatorial, as a single gene with multiple alternative exons can generate dozens or even thousands of theoretical isoforms. Modeling full-length isoform distributions from short-read sequencing data involves dealing with the uncertainty of fragment assignment, which is typically addressed through methods that estimate percent spliced in values or transcript-level quantification using expectation-maximization. Deep learning can bypass some of these limitations by predicting junction-level read counts directly from sequence and regulatory features, but this shifts the burden to generating high-resolution splicing annotations that cover rare, exercise-specific isoforms. A systems approach must therefore incorporate active learning loops in which model uncertainty is used to guide further targeted long-read sequencing, which can resolve full-length isoforms and validate predicted splice variants. This interplay between model prediction and experimental validation forms a cyber-physical feedback loop that accelerates the discovery of novel metabolic isoforms.

From a structural perspective, the integration of splicing and expression can benefit from graph-based representations where exons are nodes and splice junctions are edges, forming a splicing graph whose edge weights reflect usage frequencies that change after exercise. Graph neural networks operating on these splicing graphs can model the local constraints imposed by spliceosome assembly, such as mutual exclusivity between certain exons, without requiring the model to explicitly parameterize every possible isoform combination. This reduces the effective dimensionality of the problem and enhances biological plausibility. Combined with attention-based regulatory models that ingest local chromatin accessibility and transcription factor binding data, such graphs can capture the hierarchy from exercise-activated signaling kinases to chromatin remodeling and splicing factor phosphorylation, ultimately mapping the complete chain of metabolic adaptation.

4. Infrastructure, Data Governance, and Scalable Deployment

The promise of deep learning for exercise-induced metabolic adaptation is inextricably linked to the availability of large, diverse, and well-curated multi-omic muscle datasets. Unlike genomic sequencing projects that can aggregate DNA from millions of individuals, obtaining human skeletal muscle biopsies is invasive and resource intensive, creating a natural tension between sample depth and cohort size. This has led to a fragmented data landscape where individual laboratories generate small, richly phenotyped exercise cohorts using varying protocols for muscle sampling, RNA extraction, and sequencing depth. Building the infrastructure to harmonize these datasets into a single deep-learning-ready resource requires

a federated data architecture that respects the sovereignty of each contributing site while enabling machine learning at scale.

Federated learning, in which models are trained locally on each institutional dataset and only model gradients or parameters are aggregated centrally, provides a technically elegant solution that mitigates the need to transfer sensitive genomic and clinical data across jurisdictional boundaries. However, its application to transcriptomic deep learning presents distinct system-level hurdles. The distribution of alternative splicing patterns, expression quantitative trait loci, and even RNA-seq library preparation biases can vary systematically across sites, leading to a non-independent and non-identically distributed data environment that degrades the convergence and fairness of federated aggregation algorithms. Personalized federated learning strategies that learn a mixture of global and site-specific parameters are one avenue to address this statistical heterogeneity, but they introduce governance questions about how much local model specialization should be permitted before the central model loses generalizability across populations.

Data governance frameworks must also contend with the ethical imperative to return clinically actionable findings to participants. When a deep learning model detects a pathologic splicing variant that predisposes an individual to metabolic myopathy or impairs exercise-induced benefits, the research infrastructure must have pre-established, ethically vetted channels for clinical validation and communication. This pushes exercise genomics infrastructures toward the model of a learning health system, where research and clinical care data cycles are integrated under a unified consent and privacy architecture. Differential privacy guarantees, which mathematically limit the risk of individual re-identification even from model parameters, become essential in such environments. The calibration of the privacy budget, however, must be weighed against the potential loss in predictive accuracy for rare isoforms or population-specific splice variants, creating a governance trade-off that cannot be solved at the purely technical level.

Sustainability is another infrastructure concern. The computational demands of training large transformer-based models on multi-omic muscle data, especially when combined with privacy-preserving techniques and continuous model updates from streaming exercise study data, necessitate substantial carbon footprints and energy-optimized hardware. Green AI principles and model distillation onto efficient edge devices for deployment in wearable-integrated health coaching platforms represent important counterbalances that the field must adopt from the outset to avoid locking exercise precision medicine into an environmentally unsustainable trajectory.

5. Robustness, Fairness, and Interpretability

The deployment of deep learning models that predict metabolic adaptation from gene expression and splicing patterns raises acute concerns about robustness and fairness. Skeletal muscle biology differs between sexes, across ancestral backgrounds, and among individuals with varying metabolic disease status, and these differences manifest in both baseline transcriptome composition and exercise-induced regulatory dynamics. If a splicing prediction model is trained predominantly on data from male individuals of European descent, its accuracy in predicting isoform switches in response to exercise in African or Asian female populations may be substantially compromised, leading to inequitable performance in downstream health recommendations. This is not merely a matter of sample size balancing; the genetic architecture of splicing regulation contains population-specific expression splicing

quantitative trait loci that require explicit modeling of local haplotypic background and linkage disequilibrium patterns.

From a systems standpoint, fairness must be architected into the model pipeline rather than treated as a post-hoc remediation. This requires stratified evaluation on ancestrally and geographically defined sub-populations using an agreed-upon set of fairness metrics that capture both calibration and error rate parity across groups. It also demands the collection of population-specific functional genomics data, such as muscle-derived induced pluripotent stem cell lines that can be used to experimentally validate predicted splice variants in diverse genetic contexts. Adversarial training techniques that attempt to remove population-identifying information from latent representations can improve distributional robustness, but they carry the risk of inadvertently erasing biologically meaningful variation that is correlated with ancestry and essential for personalized adaptation. Navigating this tension requires close collaboration between machine learning engineers, statistical geneticists, and community stakeholders to define the boundaries of acceptable performance variation.

Interpretability of deep splicing models is essential not only for scientific discovery but also for establishing trust among clinicians and end-users. While SHAP and integrated gradient approaches can highlight the contribution of individual nucleotides or sequence motifs to a specific splice site prediction, the high-dimensional, interdependent nature of splicing regulation means that local feature attribution often fails to capture the emergent logic of the system. Concept-based interpretability, where the model is probed for known biological concepts such as RNA-binding protein motif activity or chromatin state, offers a more structured lens. For example, it becomes possible to show that after a bout of high-intensity interval training, the model's splicing predictions drive an isoform shift in PDK4 because the latent representation corresponding to FOXO1 activity is increased, a finding that aligns with the known insulin-sensitizing and metabolic stress response pathways. Such concept-based explanations can serve as a scientific validation tool that reinforces confidence in the model's generalizability and informs subsequent mechanistic experiments.

6. Translational and Policy Implications

Translating deep learning models of exercise-induced gene expression and splicing into real-world health interventions involves far more than achieving state-of-the-art accuracy on held-out test sets. The models must be embedded into digital health ecosystems that deliver personalized exercise prescription, monitor adaptive responses through circulating molecular markers, and adjust recommendations in an iterative, closed-loop fashion. This vision intersects with the rapid growth of consumer wearables and direct-to-consumer genetic testing, creating a landscape in which an individual's polygenic scores for metabolic traits could be combined with transcriptomic-informed deep learning models to forecast the benefits of aerobic versus resistance exercise with unprecedented granularity. However, without robust regulatory frameworks, such forecasts could be misinterpreted as deterministic predictions, undermining the autonomy-respecting communication that is essential in lifestyle counseling.

Policy governing AI-based exercise recommendations must address several unique dimensions. First, the definition of clinical validity for a splicing- or expression-based prediction tool in the context of exercise physiology remains ambiguous, as regulatory bodies have traditionally focused on diagnostic or therapeutic devices rather than wellness and optimization products. Classifying these models as low-risk wellness tools may exempt them from rigorous pre-market evaluation, yet their outputs can profoundly influence health behaviors and may even intersect with clinical decisions about medication adjustments for

metabolic diseases. A tiered regulatory model based on the intended use and potential for harm—ranging from general fitness guidance to medically prescribed exercise for type 2 diabetes management—offers a pragmatic path forward. Second, the downstream liability implications are undefined: if a deep learning model, trained on a specific population, erroneously predicts a minimal metabolic response for an individual and that individual subsequently develops worsening insulin resistance after following the AI-recommended regimen, the distribution of responsibility across software developers, data contributors, health coaches, and the individuals themselves must be clarified through updated legal and insurance frameworks.

The translational pathway also demands that deep learning models be interpretable and informative not only for bioinformaticians but for clinicians and exercise physiologists who must communicate findings to patients. Decision support interfaces that visualize isoform networks and highlight the biological rationale behind a recommended exercise modality can bridge this gap. Furthermore, as genetically encoded ionic-stress sensors that reveal protons as sleep-regulating signals [15] become integrated with human phenotypic monitoring, the possibility arises for a new breed of multi-modal deep learning systems that fuse transcriptomic data with real-time physiological ionic flux measurements. Policy frameworks must proactively address the dual-use nature of such rich physiological-molecular data, ensuring that it is not repurposed for discriminatory health insurance underwriting or coercive wellness programs that penalize individuals for genetic or molecular profiles beyond their control.

7. Conclusion

Deep learning is poised to fundamentally reshape our understanding of how skeletal muscle adapts to exercise at the gene expression and splicing levels, offering a systemic lens that spans from nucleotide sequence to organismal metabolic phenotype. Realizing this potential, however, requires a deliberate and integrative approach that marries architectural innovation in sequence-based neural networks with robust infrastructure for federated, privacy-safe multi-omic data sharing. The structural trade-offs between convolutional efficiency and attention-based long-range modeling, between joint expression-splicing tasks and performance stability, and between model accuracy and demographic fairness represent critical design decisions that will determine the ultimate utility of these systems in diverse populations. Beyond the algorithm, the governance frameworks for exercise genomics AI must mature in step with the technology, embedding fairness, interpretability, and sustainability into the deployment cycle from the earliest stages. As novel molecular sensors illuminate the ionic stress dimension of muscle metabolism and its systemic linkages to sleep and energy balance, the next generation of deep learning models will face the challenge of integrating these disparate physiological layers into a coherent, personalized forecast of metabolic adaptation. Meeting this challenge will require sustained collaboration across computational sciences, muscle physiology, bioethics, and health policy, but the promise of democratizing precision exercise medicine for the prevention and management of metabolic disease stakes a compelling claim on our collective efforts.

References

1. Pillon, N. J., Gabriel, B. M., Dollet, L., Smith, J. A. B., Sardón Puig, L., Botella, J., ... & Zierath, J. R. (2020). Transcriptomic profiling of skeletal muscle adaptations to exercise in humans and rodents. bioRxiv. https://doi.org/10.1101/2020.06.15.151555

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

3. Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J. F., Darbandi, S. F., Knowles, D., Li, Y. I., ... & Merker, J. (2019). Predicting splicing from primary sequence with deep learning. Cell, 176(3), 535–548.

4. Wang, W., Liew, W. L., Huang, S., Chan, E., Tan, A. L. M., Tian, C., ... & Liu, B. (2025). Impact of polymorphisms on gene expression and splicing in response to exercise and diet-induced weight loss in human skeletal muscle tissues. Cell Genomics, 5(9).

5. Hoppeler, H., Baum, O., Lurman, G., & Mueller, M. (2011). Molecular mechanisms of muscle plasticity with exercise. Comprehensive Physiology, 1(3), 1383–1412.

6. Ruegsegger, G. N., & Booth, F. W. (2018). Health benefits of exercise. Cold Spring Harbor Perspectives in Medicine, 8(7), a029694.

7. Zhou, J., & Troyanskaya, O. G. (2015). Predicting effects of noncoding variants with deep learning–based sequence model. Nature Methods, 12(10), 931–934.

8. Avsec, Ž., Agarwal, V., Visentin, D., Ledsam, J. R., Grabska-Barwinska, A., Taylor, K. R., ... & Kelley, D. R. (2021). Effective gene expression prediction from sequence by integrating long-range interactions. Nature Methods, 18(10), 1196–1203.

9. Ule, J., & Blencowe, B. J. (2019). Alternative splicing regulatory networks: functions, mechanisms, and evolution. Molecular Cell, 76(2), 329–345.

10. Rappoport, N., & Shamir, R. (2018). Multi-omic and multi-view clustering algorithms: review and cancer benchmark. Nucleic Acids Research, 46(20), 10546–10562.

11. GTEx Consortium. (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science, 369(6509), 1318–1330.

12. Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3, 160018.

13. Price, W. N., & Cohen, I. G. (2019). Privacy in the age of medical big data. Nature Medicine, 25(1), 37–43.

14. Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., ... & Cardoso, M. J. (2020). The future of digital health with federated learning. NPJ Digital Medicine, 3, 119.

15. Ji, Z., Liu, J., Wang, B., Wei, S., Bian, Y., Zeng, W., ... & Ma, D. K. (2026). A genetically encoded ionic-stress sensor reveals protons as a sleep driver. bioRxiv, 2026-01.

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

17. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447–453.

18. Koh, P. W., & Liang, P. (2017). Understanding black-box predictions via influence functions. Proceedings of the 34th International Conference on Machine Learning, 70, 1885–1894.

19. Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). Deep learning with differential privacy. Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, 308–318.

20. Finlayson, S. G., Bowers, J. D., Ito, J., Zittrain, J. L., Beam, A. L., & Kohane, I. S. (2019). Adversarial attacks on medical machine learning. Science, 363(6433), 1287–1289.

21. Sartor, F., & Vernillo, G. (2023). Sleep and exercise: a reciprocal relationship. Sports Medicine, 53(Suppl 1), 61–72.

22. Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine, 25(1), 44–56.

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

24. Schwartz, R., Vassilev, A., Greene, K., Perine, L., Burt, A., & Hall, P. (2022). Towards a standard for identifying and managing bias in artificial intelligence (NIST Special Publication 1270). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.1270

25. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115–118.