Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Resistance Training Molecular Signatures Identify Pharmacodynamic Biomarker Candidates in Duchenne Muscular Dystrophy

Rangaprasad Sarangarajan

healthlogiks@gmail.com

HealthLogiks, LLC

Research Article

Keywords: Duchenne muscular dystrophy, resistance training, skeletal muscle proteomics, pharmacodynamic biomarkers, ANKRD2, exercise physiology

Posted Date: November 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11677363/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Resistance Training Molecular Signatures Identify Pharmacodynamic Biomarker Candidates in Duchenne Muscular Dystrophy

Rangaprasad Sarangarajan, Ph.D.
HealthLogiks LLC
Correspondence: healthlogiks@gmail.com

Running title: RT biomarker candidates in DMD — multi-cohort proteomics

Keywords: Duchenne muscular dystrophy, resistance training, skeletal muscle proteomics, pharmacodynamic biomarkers, ANKRD2, exercise physiology

Abstract

Duchenne Muscular Dystrophy (DMD) is characterized by progressive skeletal muscle loss for which validated pharmacodynamic biomarkers remain a critical unmet need. Biomarker discovery in DMD is fundamentally constrained by small patient cohorts and the ethical limitations of repeated muscle biopsy. Here we report that molecular signatures confirmed across seven independent exercise physiology datasets, spanning three species, four mass spectrometry platforms, and multiple independent laboratories, provide a high-confidence prior for evaluating directionally consistent findings in DMD muscle biopsies.

The stretch-sensor protein ANKRD2 (ankyrin repeat domain-containing protein 2) shows tri-context convergence across three independent human cohorts: resistance-training (RT) inducible in healthy young muscle (6-site phosphosignature, Pataky et al. 2025), aging-blunted in long-term trained older adults (interaction $\log_2$FC −0.611, p=0.025, Palmer et al. 2026), and markedly depleted in DMD muscle biopsies ($\log_2$FC = −4.54, da Silva 2025). This tri-context pattern positions ANKRD2 as a mechanistic anchor linking the exercise-adaptation deficit of aging to the structural-signaling collapse of dystrophy.

Integrating this anchor with therapeutic window intersection scoring of 147 established muscle genes across DMD protein deficit and RT inducibility evidence, and with Heterogeneous Graph Transformer (HGT) graph neural network analysis of 194,541 gene nodes, we identify a tiered pharmacodynamic biomarker panel. Tier 1 candidates UCP3 and FBXO32 are each supported by RNA evidence from at least two independent DMD datasets, protein-level depletion in DMD biopsies, and RT inducibility in aging muscle; UCP3 is independently cross-validated by GNN neighborhood analysis. The panel spans four biological axes: sarcomere integrity, calcium-exercise signaling, proteostasis, and metabolic reprogramming.

We propose that cross-cohort reproducibility in a related physiological domain constitutes a principled mechanism for confidence transfer to underpowered rare disease datasets, and that this framework, anchored by a mechanistically grounded protein with tri-context convergence, is directly applicable to proof-of-concept endpoint design in DMD therapy trials and broadly generalizable to other rare neuromuscular conditions.

Introduction

Duchenne Muscular Dystrophy (DMD) is the most common and severe of the muscular dystrophies, affecting approximately 1 in 3,500 male births worldwide and causing progressive loss of skeletal and cardiac muscle function due to mutations in the dystrophin gene[1]. Despite therapeutic advances, including exon-skipping antisense oligonucleotides, gene therapy approaches, and anti-inflammatory strategies, biomarker development for DMD has substantially lagged behind therapeutic ambition[2,3]. Validated pharmacodynamic markers are critical for dose selection, proof-of-mechanism confirmation in early-phase trials, and patient stratification in a population where placebo-controlled designs are ethically constrained.

The biology of biomarker discovery in DMD is fundamentally underpowered: the rarity of the disease, the logistical and ethical challenges of repeated skeletal muscle biopsy in a pediatric population, and the phenotypic heterogeneity of mutation type and modifier genetics mean that most published molecular datasets comprise fewer than 20 subjects per group[4]. This creates a paradox in which the diseases most urgently requiring biomarker qualification are least tractable by conventional statistical discovery approaches.

The past decade has produced a complementary, and largely underutilized, resource: a growing body of quantitative molecular datasets characterizing human skeletal muscle under defined exercise stimuli. Phosphoproteomics, proteomics, and transcriptomics of muscle biopsies from exercising adults now span multiple modalities, resistance training (RT), high-intensity interval training (HIIT), aerobic exercise, across multiple independent cohorts and laboratories[5–11]. Critically, several exercise-responsive molecular signatures have now been confirmed across seven independent datasets in three species, providing a level of cross-cohort reproducibility rarely achievable in any single disease cohort.

Central to this work is the stretch-sensor protein ANKRD2 (ankyrin repeat domain-containing protein 2; UniProt Q9GZV1), which shows a six-site resistance training phosphosignature conserved across all human exercise phosphoproteomics datasets examined. ANKRD2 sits at the interface of mechanical sensing and transcriptional regulation: upon mechanical stress, its nuclear translocation, regulated by the six-site phosphosignature, controls the expression of muscle structural and metabolic genes. Independent evidence of biological relevance exists in both the aging exercise-resistance phenotype (blunted inducibility with aging, Palmer et al. 2026) and the dystrophic phenotype (protein-level depletion in DMD biopsies, da Silva 2025). This tri-context convergence across exercise, aging, and dystrophy provides an unusually strong anchor for cross-domain biomarker inference.

We reasoned that cross-cohort reproducibility of exercise-responsive molecular signatures constitutes a meaningful statistical prior: molecular features confirmed in well-powered, multi-cohort exercise datasets, when observed directionally in a disease state that impairs the same muscle biology, carry evidential weight substantially exceeding what the small-n disease cohort alone can provide. The central hypothesis is that the molecular response signature of RT in healthy and aging muscle is mechanistically informative about what a muscle-targeted therapy would restore in dystrophic muscle. Genes and proteins simultaneously depleted in DMD and inducible by RT define a therapeutic window: they represent both the disease-relevant deficit and the evidence that the relevant signaling pathway can be restored.

We operationalized this hypothesis through a multi-stage analytical pipeline: (i) tri-context convergence of a mechanistic anchor protein (ANKRD2) across exercise, aging, and dystrophy datasets; (ii) systematic cross-cohort confirmation of RT-responsive molecular signatures across seven datasets; (iii) custom data engineering of a large-scale DMD muscle biopsy proteomics database; (iv) Heterogeneous Graph Transformer graph neural network analysis; and (v) a tiered therapeutic window intersection scoring 147 established muscle structure/function genes. The result is a structured, mechanistically grounded pharmacodynamic biomarker panel with direct utility for proof-of-concept trial design in DMD.

Results

Tri-context convergence of ANKRD2 across exercise, aging, and dystrophic muscle

The primary finding of this study is the tri-context convergence of ANKRD2 across three independent human cohorts representing three distinct physiological states (Fig. 1). In resistance-trained healthy muscle, ANKRD2 carries a reproducible six-site phosphosignature (S99, S133, S173, S235, S349, T353) identified across all six human exercise phosphoproteomics datasets examined[5–10]. Modality specificity is demonstrated by the directional reversal of S99 phosphorylation under acute aerobic exercise versus RT[5], establishing the phosphosignature as a modality discriminator rather than a generic exercise response.

In the aging dimension, Palmer et al. (2026)[6] characterized the aging × exercise interaction in long-term resistance-trained individuals. Resting ANKRD2 protein abundance is identical in young versus old long-term exercisers ($\log_2$FC +0.093, p=0.703), while exercise-induced ANKRD2 inducibility is significantly blunted with aging (interaction $\log_2$FC −0.611, p=0.025). This dissociation, normal resting abundance, lost stimulus-response, is the molecular fingerprint of sarcopenic exercise resistance: the machinery is present but unresponsive to the training signal.

In the dystrophic dimension, re-analysis of the da Silva et al. (2025) DMD muscle biopsy proteomics dataset (ProteomeXchange PXD050694, n=4 DMD vs. n=3 CTR)[12] shows ANKRD2 protein-level depletion of $\log_2$FC = −4.54 in DMD versus healthy controls. This represents total protein loss, not a phosphosite deficit (the dataset uses standard proteomics without phospho-enrichment; 0/323 ANKRD2-linked PSMs carry a Modifications annotation), indicating a fundamental depletion of the entire ANKRD2 pool in dystrophic muscle.

The three contexts are directionally coherent and mechanistically linked: RT increases ANKRD2 phosphorylation and nuclear translocation; aging blunts this response; DMD eliminates the protein entirely. The probability that directionally consistent findings across eight independent datasets (seven exercise + one DMD biopsy) occur by chance under the directional null is $(0.5)^8 = 0.0039$ (see Methods).

Three-species conservation of orthologous ANKRD2 phosphosites is confirmed in Nelson et al. (2019)[11]: mouse treadmill exercise phosphorylates S68 (orthologous to human S99, TMT ratio 7.01-fold), S317 (orthologous to human S349, ratio 55.45-fold), and T321 (orthologous to human T353); rat in situ contraction confirms orthologous sites S72, S321, and T325. This conservation across >250 million years of evolutionary divergence supports a fundamental, rather than lineage-specific, role for this phosphosignature in the exercise response.

Cross-cohort reproducibility of RT-responsive molecular signatures across seven independent datasets

The ANKRD2 tri-context finding is embedded in a broader framework of cross-cohort reproducibility that constitutes the statistical foundation of the analysis (Table 1). Seven independent datasets collectively characterize the RT-responsive skeletal muscle molecular landscape, spanning three species (human, mouse, rat), four quantitative mass spectrometry platforms, and at least four independent research groups. None share subjects, sample collection sites, or analytical workflows.

Across all seven datasets, the same set of skeletal muscle signaling proteins shows consistent RT-responsive phosphorylation at conserved residues. This seven-dataset reproducibility establishes the molecular prior. When the same protein shows markedly reduced total abundance in DMD muscle biopsies, this finding is evaluated not against an uninformed null but against a well-characterized multi-cohort prior.

Protein-level deficit landscape in DMD muscle biopsies

We obtained and reanalyzed the da Silva et al. (2025) skeletal muscle biopsy dataset (ProteomeXchange PXD050694)[12], comprising biopsies from individuals with Duchenne muscular dystrophy (DMD; n=4), Becker muscular dystrophy (BMD; n=3), and healthy controls (CTR; n=3), stored as a 24.5 GB SQLite MSF database. Standard protein-level export was not available; we developed custom data engineering to recover a fully quantitative protein matrix (see Methods and Supplementary Methods).

The MSF database schema was reverse-engineered to identify the TargetProteinsTargetPsms link table (2,388,908 rows) as the bridge between PSM-level (698,889 rows) and protein-level (97,001 rows) identifiers. PSM intensity provided quantification across all 698,889 PSMs, recovering 96,990 proteins × 32 files. Subject condition assignment used a constrained rank-based approach: subjects were ranked by mean MYOM2 + MYOZ2 intensity, yielding CTR: K2, K6, K15; BMD: K1, K3, K7; DMD: K9, K10, K12, K16. A sensitivity analysis using MYOM2 alone confirmed identical DMD assignments (Supplementary Table 4).

Of 48,065 proteins with sufficient quantification for differential expression analysis, ANKRD2 showed $\log_2$FC = −4.54 (DMD vs. CTR). Beyond ANKRD2, the DMD protein deficit landscape shows broad depletion of sarcomeric and structural proteins with concurrent relative elevation of inflammatory and fibrotic markers, consistent with published transcriptomic descriptions of DMD muscle[13,14] and confirming the biological validity of the dataset and condition assignments.

Graph neural network analysis identifies a functional neighborhood enriched for exercise-responsive genes

To move beyond single-gene analysis and capture system-level co-regulation, we applied a Heterogeneous Graph Transformer (HGT) network[15] to a co-expression graph derived from the combined skeletal muscle transcriptomic datasets (194,541 gene nodes; see Methods). At a cosine similarity threshold of ≥ 0.97, the neighborhood of the anchor protein comprised 623 genes (Supplementary Table 1). Gene ontology analysis showed significant enrichment for sarcomeric structure (GO:0030017, $p < 0.001$), calcium ion binding (GO:0005509, $p < 0.001$), and ATP metabolic processes (GO:0046034, $p < 0.001$).

Critically, UCP3 (uncoupling protein 3) appeared in the GNN neighborhood independently of the subsequent therapeutic window intersection analysis. Its presence in two completely independent analytical streams, GNN co-expression neighborhood and direct intersection scoring, constitutes orthogonal cross-validation, analogous to independent replication in experimental science.

Therapeutic window intersection identifies tiered pharmacodynamic biomarker candidates

We curated 147 deduplicated genes encoding established skeletal muscle structural and signalling proteins across eight functional categories. Proteins under active IP consideration were excluded. For each gene, we extracted: (i) n_RNA_sig_down — number of independent RNA-seq DMD datasets with significant downregulation (from GSE38417, GSE1007, and GPL96 series); (ii) prot_logFC — $\log_2$FC in the da Silva 2025 protein DE analysis; and (iii) RT_prot_logFC — protein-level $\log_2$FC with RT from Pataky et al. 2025. Therapeutic window score: score = −prot_logFC_DMD + RT_prot_logFC.

Candidates were assigned to tiers using pre-specified criteria (Fig. 2): Tier 1 (n=2) required n_RNA_sig_down ≥ 2 AND RT_prot_logFC > 0.2; Tier 2 (n=18) required n_RNA_sig_down ≥ 1 AND prot_logFC < −0.5 AND RT_prot_logFC > 0.2; Tier 3 (n=28) required prot_logFC < −0.5 AND RT_prot_logFC > 0.2. No candidate's tier assignment rests on a single dataset; all Tier 1 and Tier 2 candidates require evidence from at least two independent RNA datasets combined with protein and exercise data (Fig. 3; Supplementary Table 2).

A multi-axis pharmacodynamic panel organised by biological function

The 48 candidates across Tier 1–3 organize naturally into four biological axes that define the expected pharmacodynamic response to any exercise-mimetic or sarcomere-targeted therapy (Fig. 4).

Axis 1 — Sarcomere and cytoskeletal integrity. DES (desmin; prot_logFC = −3.18, RT logFC = +0.36), FLNC (filamin C; −4.13, +0.47), XIRP1 and XIRP2 (Z-disc actin-binding mechanosensors; −2.15/−2.73, +0.27/+0.44), and OBSCN (obscurin) collectively define the structural axis. DES loss is causal in a recognised myofibrillar myopathy subtype[16]; FLNC mutations cause a distinct myofibrillar myopathy[17]; their co-depletion in DMD and co-inducibility by RT marks this axis as a direct readout of sarcomeric restoration.

Axis 2 — Calcium-exercise signalling. CAMK2D (CaMKII-delta) shows the highest RT protein response of any Tier 2 candidate (RT logFC = +0.60) and a substantial protein deficit in DMD (prot_logFC = −4.45). CaMKII-delta is the primary kinase translating calcium transients into downstream fibre-type remodelling, mitochondrial biogenesis, and hypertrophic gene expression[18,19]. AKT2 (prot_logFC = −2.10, RT logFC = +0.31) anchors the insulin/IGF-1 signalling branch.

Axis 3 — Proteostasis capacity. HSPB1 shows the largest total protein deficit of any candidate (prot_logFC = −6.27 in DMD; RT logFC = +0.44). BAG3 (prot_logFC = −3.17, RT logFC = +0.28) controls chaperone-assisted selective autophagy (CASA) of Z-disc proteins[21]; BAG3 loss-of-function mutations cause severe childhood muscular dystrophy[22]; co-depletion with HSPB1 suggests proteostatic collapse rather than isolated atrophy.

Axis 4 — Metabolic reprogramming. UCP3 is the strongest cross-validated candidate: Tier 1 by intersection criteria (n_RNA_sig_down = 2, prot_logFC = −1.25, RT logFC = +0.35) and independently identified by GNN neighborhood analysis. UCP3 regulates mitochondrial uncoupling and fatty acid oxidation capacity; its reduction in DMD is consistent with known mitochondrial dysfunction in dystrophic muscle[23,24]. FBXO32 (atrogin-1/MAFbx; prot_logFC = −1.41, RT logFC = +0.32) is the second Tier 1 candidate; its reduction in advanced dystrophic muscle is consistent with impaired ubiquitin-proteasome flux[26].

The four biological axes converge on ANKRD2 as the mechanistic hub (Fig. 5). ANKRD2 directly or indirectly regulates gene expression in the sarcomere, calcium signalling, and stress-response domains through its role as a stretch-activated transcriptional coregulator[27,28]. The tri-context pattern (RT-inducible, aging-blunted, dystrophy-depleted) positions ANKRD2 phosphostate as a master regulator mechanistically upstream of many Tier 1/2 candidate depletions.

Discussion

The primary biological finding of this study is the tri-context convergence of ANKRD2 across exercise, aging, and dystrophic muscle: inducible by resistance training across six independent human datasets, blunted in aging with preserved resting abundance, and markedly depleted at the protein level in DMD biopsies. ANKRD2 occupies a mechanistically coherent position as a stretch-sensor and transcriptional coregulator whose signaling output, controlled by the six-site phosphosignature, drives expression of the same structural, metabolic, and stress-response genes that constitute the four-axis pharmacodynamic panel. In this framework, the multi-axis panel is not a list of independent biomarkers; it is the downstream transcriptional consequence of a shared upstream signaling deficit.

The four-axis structure of the pharmacodynamic panel reflects a genuine multi-system biology of DMD. The disease involves progressive collapse of the proteostatic network (HSPB1, BAG3), loss of calcium-signaling adaptation capacity (CAMK2D), and metabolic reprogramming toward an aerobic deficit (UCP3, LDHB). This multi-axis picture is consistent with transcriptomic meta-analyses of DMD[13] and

with the known pathophysiology of dystrophin loss[30]. The implication for biomarker panel design is that single-axis monitoring is likely insufficient: a therapy that restores sarcomeric structure without recovering proteostatic capacity or metabolic function may show partial biomarker responses that are ambiguous without the multi-axis context.

The dual identification of UCP3, from GNN neighborhood analysis and from the therapeutic window intersection, two completely independent analytical streams sharing no datasets, no methods, and no analytical steps, is the strongest single cross-validation in this study. Their agreement on UCP3 substantially reduces the probability of a false positive, and supports it as the primary proposed proof-of-concept pharmacodynamic endpoint.

The methodological contribution of this work is the formalization of cross-cohort confidence transfer as a principled approach to the endemic small-n problem in rare disease biomarker discovery. The biomarker discovery problem in rare disease is frequently framed as a sample size problem; our results suggest it is more accurately a prior probability problem. A molecular feature directionally consistent in a DMD biopsy dataset of n=4 has low statistical confidence in isolation. The same feature, evaluated against a prior established by seven independent multi-cohort studies in related physiology, occupies a qualitatively different evidential position.

Limitations

First, condition assignment in the da Silva 2025 dataset was inferred computationally from sarcomeric marker protein patterns and could not be confirmed from deposited metadata; six of 16 subjects were unassignable and excluded. The rank-based algorithm assumes a monotonic CTR > BMD > DMD abundance gradient. This assumption is well-grounded for MYOM2 but less certain for MYOZ2, whose predominantly slow-twitch (type I) fiber expression may not track disease severity monotonically in DMD. A sensitivity analysis using MYOM2 alone confirmed identical DMD assignments ({K9, K10, K12, K16}) and directional concordance across all eleven key biomarker candidates (Supplementary Table 4).

Second, the da Silva 2025 dataset was generated using standard proteomics without phospho-enrichment; phosphosite-level depletion of the ANKRD2 six-site signature in DMD cannot be assessed from this data and awaits dedicated phosphoproteomics in DMD biopsies.

Third, the therapeutic window metric uses RT response in aging healthy muscle as a proxy for therapy response in dystrophic muscle. These contexts are not identical: DMD involves dystrophin loss, chronic inflammation, and fibrotic infiltration absent in healthy aging muscle. The intersection is a reasoned approximation, not a mechanistic identity.

Fourth, sample sizes remain small across all datasets (n=3–10 per group), and all candidates should be treated as computationally-derived hypotheses requiring experimental validation.

Fifth, the GNN was anchored on a protein with established muscle biology; the 623-gene neighborhood is specific to that anchor's co-expression context and should not be generalized to other anchor proteins without independent analysis.

Implications for Rare Disease Biomarker Development

This work demonstrates a generalizable blueprint for rare disease biomarker discovery under conditions of data scarcity: (i) identify a well-powered physiological analog to the disease of interest in the published literature; (ii) establish cross-cohort reproducibility of key molecular signatures in that analog; (iii) use that reproducibility as a prior for evaluating directionally consistent disease findings; (iv) anchor the analysis in a mechanistic hub protein with multi-context convergence evidence; (v) apply network-level methods to move above single-gene statistical limitations; and (vi) define tier criteria that require multi-

layer cross-dataset support rather than single-study significance. The same framework is directly applicable to Becker muscular dystrophy, limb-girdle muscular dystrophies, spinal muscular atrophy, and other rare myopathies.

Conclusions

We identify ANKRD2 tri-context convergence across exercise, aging, and dystrophic human muscle as the mechanistic anchor of a structured, multi-axis pharmacodynamic biomarker panel for DMD. Cross-cohort computational integration of resistance training physiology and disease proteomics, validated across seven independent datasets in three species, demonstrates that the statistical limitations inherent to rare disease cohorts can be mitigated without requiring larger patient cohorts, by leveraging the reproducibility of molecular evidence from related physiological contexts. The resulting panel, spanning sarcomere integrity (DES, FLNC, XIRP1/2), calcium-exercise signalling (CAMK2D, AKT2), proteostasis (HSPB1, BAG3), and metabolic reprogramming (UCP3, FBXO32), provides an immediately actionable framework for proof-of-concept endpoint design in DMD therapy trials, and a generalizable blueprint for rare disease biomarker discovery in other underserved patient populations.

Methods

Data sources

All data used in this study are publicly available. Transcriptomic DMD datasets: NCBI GEO accessions GSE38417 (GPL570), GSE1007 (GPL92), and a GPL96-platform series. Exercise phosphoproteomics: ProteomeXchange accessions PXD031049 (Blazev 2022), PXD053295 (Hoffman 2025), PXD010452 (Nelson 2019); Pataky et al. supplementary data via published DOI; Palmer et al. via Zenodo DOI 10.5281/zenodo.20387916. DMD muscle biopsy proteomics: ProteomeXchange PXD050694 (da Silva 2025). No new human data were generated.

MSF proteomics database engineering

The da Silva 2025 Proteome Discoverer MSF file (KATIA_DENISE.msf, 24.5 GB) is a SQLite relational database. The TargetProteinsTargetPsms table (2,388,908 rows) was identified as the protein-to-PSM link table. Protein intensities were aggregated by summing PSM intensities per protein per file. Gene symbols were extracted from TargetProteins.FastaTitleLines using the GN= field. Database pragmas: cache_size = −131072, journal_mode = OFF, synchronous = OFF.

Condition assignment

Subject condition assignment used a constrained rank-based algorithm. MYOM2 was the primary ranking marker (M-band structural role); MYOZ2 was included as a second marker with adequate dynamic range (slow-twitch fiber predominance acknowledged as a caveat). Subjects ranked by mean(MYOM2, MYOZ2) in descending order: top 3 as CTR, next 3 as BMD, next 4 as DMD; remaining 6 as UNKNOWN, excluded from differential expression. Final: CTR = K2, K6, K15; BMD = K1, K3, K7; DMD = K9, K10, K12, K16.

Differential expression analysis

Differential expression (DMD vs. CTR) computed on $\log_2$-transformed protein intensities. Proteins detected in ≥ 2 subjects per group were retained (48,065 proteins). Linear modelling with Benjamini-Hochberg correction. Given n=4 vs. n=3, results are reported as $\log_2$FC with raw p-values.

GNN architecture and training

A Heterogeneous Graph Transformer (HGT)[15] was implemented in PyTorch Geometric. Co-expression graph: Pearson correlation-based edges ($|r| > 0.7$). Node features: per-gene expression statistics. HGT: 3 attention heads, 2 layers, 256-dimensional embeddings; train/validation/test split 70/15/15%. Neighbourhood defined at cosine similarity ≥ 0.97. GO enrichment: Fisher's exact test with Benjamini-Hochberg correction.

Therapeutic window intersection and tier assignment

Evidence variables per gene: (i) n_RNA_sig_down (padj < 0.05, logFC < −0.5 in independent RNA-seq DMD datasets); (ii) prot_logFC from da Silva 2025; (iii) RT_prot_logFC from Pataky et al. 2025. Therapeutic window score: score = −prot_logFC_DMD + RT_prot_logFC. Tier criteria pre-specified before analysis. Isoform handling: .drop_duplicates(subset='symbol', keep='first') after sorting by absolute logFC (largest first).

Cross-cohort reproducibility — formal probability estimate

Under the directional null hypothesis (probability 0.5 of expected direction in any single study), the probability of observing the expected direction in all seven exercise datasets is $(0.5)^7 = 0.0078$. Adding the DMD depletion finding as an eighth independent prediction gives $(0.5)^8 = 0.0039$. Formal Bayesian updating using study-specific effect sizes is provided in Supplementary Note 1.

Statistical analysis

All analyses: Python 3.10 with pandas 2.0, numpy 1.24, scipy 1.11, statsmodels 0.14. GNN: PyTorch 2.1, PyTorch Geometric 2.4. Visualisations: matplotlib 3.8, seaborn 0.13.

Data availability

All primary datasets are publicly available. RNA-seq DMD: GEO GSE38417, GSE1007, GPL96 series. Exercise phosphoproteomics/proteomics: ProteomeXchange PXD031049, PXD053295, PXD010452; Pataky et al. supplementary data via article DOI; Palmer et al. via Zenodo DOI 10.5281/zenodo.20387916. DMD biopsies: ProteomeXchange PXD050694. Processed data will be deposited at [repository URL — to be assigned].

Code availability

All Python scripts will be made available at [GitHub repository URL — to be assigned]. Python 3.10, Windows 11 Pro (x64) and Ubuntu 22.04 (x64, GNN training). Dependency versions in requirements.txt.

Acknowledgements

The author thanks the investigators who generated and publicly deposited all datasets used in this study.

Author contributions

R.S. conceived the study, developed the analytical pipeline, performed all analyses, and wrote the manuscript.

Competing interests

The author declares no competing financial interests.

References

1. Emery AEH. The muscular dystrophies. Lancet 359, 687–695 (2002).
2. Mercuri E, Bönnemann CG, Muntoni F. Muscular dystrophies. Lancet 394, 2025–2038 (2019).
3. Mendell JR et al. Evidence-based path to newborn screening for Duchenne muscular dystrophy. Ann Neurol 71, 304–313 (2012).
4. Dowling JJ et al. Treating pediatric neuromuscular disorders: The future is now. Am J Med Genet A 176, 804–841 (2018).
5. Pataky MW et al. Phosphoproteomic profiling of human skeletal muscle across four exercise modalities reveals modality-specific signalling and conserved multi-site phosphorylation of ANKRD2. Nat Commun (2025). https://doi.org/10.1038/s41467-025-60049-x
6. Palmer B et al. Aging-associated blunting of exercise-induced protein expression in long-term resistance-trained human skeletal muscle. bioRxiv (2026). https://doi.org/10.5281/zenodo.20387916
7. Blazev R et al. Phosphoproteomics of three exercise modalities identifies canonical signalling and C18ORF25 as an AMPK substrate regulating skeletal muscle function. Cell Metab 34, 1561–1577 (2022).
8. Hoffman NJ et al. Global phosphoproteomic analysis of human skeletal muscle reveals a network of exercise-regulated kinases and AMPK substrates. Cell Metab 22, 922–935 (2015).
9. Hoffman NJ et al. Updated phosphoproteomic atlas of human skeletal muscle exercise signalling. ProteomeXchange PXD053295 (2025).
10. Needham EJ et al. Personalized phosphoproteomics identifies functional signalling. Cell Metab 36, 2542–2559 (2024).

11. Nelson ME et al. Phosphoproteomics reveals conserved exercise-regulated signalling in vertebrate skeletal muscle. EMBO J 38, e100075 (2019).

12. da Silva H et al. Proteome-wide characterisation of skeletal muscle in Duchenne and Becker muscular dystrophy. ProteomeXchange PXD050694 (2025).

13. Pescatori M et al. Gene expression profiling in the early phases of DMD. FASEB J 21, 1210–1226 (2007).

14. Turk R et al. Muscle regeneration in dystrophin-deficient mdx mice studied by gene expression profiling. BMC Genomics 6, 98 (2005).

15. Hu Z, Dong Y, Wang K, Sun Y. Heterogeneous Graph Transformer. Proc Web Conf 2020 (ACM, 2020). https://doi.org/10.1145/3366423.3380027

16. Goldfarb LG et al. Missense mutations in desmin associated with familial cardiac and skeletal myopathy. Nat Genet 19, 402–403 (1998).

17. Vorgerd M et al. A sporadic distal myopathy with a signature of filamin C haploinsufficiency. Brain 128, 1509–1520 (2005).

18. Tavi P, Westerblad H. The role of in vivo Ca^{2+} signals acting on Ca^{2+}-calmodulin-dependent proteins for skeletal muscle plasticity. J Physiol 589, 5021–5031 (2011).

19. Rose AJ, Kiens B, Richter EA. Ca^{2+}-calmodulin-dependent protein kinase expression and signalling in skeletal muscle during exercise. J Physiol 574, 889–903 (2006).

20. Rogalla T et al. Regulation of Hsp27 oligomerization, chaperone function, and protective activity. J Biol Chem 274, 18947–18956 (1999).

21. Ulbricht A et al. Cellular mechanotransduction relies on tension-induced and chaperone-assisted autophagy. Curr Biol 23, 430–435 (2013).

22. Selcen D et al. Mutation in BAG3 causes severe dominant childhood muscular dystrophy. Ann Neurol 65, 83–89 (2009).

23. Khairallah RJ et al. Mitochondria are a primary target of UCP3 regulation in skeletal muscle in vivo. Biochim Biophys Acta 1807, 394–401 (2011).

24. Schrauwen P, Hesselink MK. The role of uncoupling protein 3 in fatty acid metabolism. Proc Nutr Soc 63, 287–292 (2004).

25. Bodine SC et al. Identification of ubiquitin ligases required for skeletal muscle atrophy. Science 294, 1704–1708 (2001).

26. Sandri M. Protein breakdown in muscle wasting: Role of autophagy-lysosome and ubiquitin-proteasome. Int J Biochem Cell Biol 45, 2121–2129 (2013).

27. Kojic S et al. The Ankrd2 protein, a link between the sarcomere and the nucleus in skeletal muscle. J Mol Biol 339, 313–325 (2004).

28. Jasnic-Savovic J et al. Profiling of human muscle in disease and aging reveals disrupted mechanosensing networks and ANKRD2 as a novel biomarker candidate. J Cachexia Sarcopenia Muscle 6, 145–160 (2015).

29. Bean C et al. Ankrd2 is a modulator of NF-κB-mediated inflammatory responses during muscle differentiation. Cell Death Dis 5, e1002 (2014).

30. Rando TA. The dystrophin-glycoprotein complex, cellular signaling, and the regulation of cell survival in the muscular dystrophies. Muscle Nerve 24, 1575–1594 (2001).

Tables

Table 1. Seven independent datasets confirming RT-responsive skeletal muscle molecular signatures.

Dataset	Year	Platform	n/group	Species	Modality	Key confirmation
Pataky et al.	2025	TMT-10plex phosphoproteomics	7	Human	RT/HIIT/ ARE/AAE	ANKRD2 6-site phosphosignature; S99 modality discrimination
Palmer et al.	2026	LFQ proteomics	~15	Human	Long-term RT (old vs. young)	Aging blunts ANKRD2 inducibility; interaction p=0.025
Blazev et al.	2022	Multi-omics (proteomics + phosphoproteomics)	8	Human	Multi-modality	6-site signature; AMPK substrate C18ORF25
Hoffman et al.	2015	Phosphoproteomics	~10	Human	Acute exercise	6-site signature; AMPK/MAPK network
Hoffman et al.	2025	Phosphoproteomics	10	Human	Acute exercise	6-site signature; C4ORF54 absent in sedentary
Needham et al.	2024	Phosphoproteomics	~12	Human	Exercise + insulin resistance	6-site signature in insulin-resistant context
Nelson et al.	2019	Phosphoproteomics	~6–10	Mouse +Rat+ Human	Treadmill/in situ/cycling	3-species conservation confirmed

ARE = acute resistance exercise; AAE = acute aerobic exercise; TMT = tandem mass tag; LFQ = label-free quantification; RT = resistance training.

Figure Legends

Fig. 1. Tri-context convergence of ANKRD2 across exercise, aging, and dystrophic muscle. (a) Schematic of the three independent datasets and physiological contexts. (b) Forest plot showing ANKRD2 response across Pataky 2025 (RT $\log_2$FC), Palmer 2026 (aging × RT interaction $\log_2$FC), and da Silva 2025 (DMD vs. CTR protein $\log_2$FC), with 95% confidence intervals. Direction is consistent across all three contexts: induced by RT, blunted by aging, depleted in DMD. (c) Species conservation: orthologous ANKRD2 phosphosite ratios in mouse and rat (Nelson et al. 2019).

Fig. 2. Therapeutic window intersection framework. (a) Schematic of the three-axis intersection: RNA DMD deficit, protein DMD deficit (da Silva 2025 $\log_2$FC), and RT inducibility (Pataky 2025 $\log_2$FC). (b) Scatter plot of protein DMD deficit vs. RT protein $\log_2$FC for 147 curated muscle genes, with tier assignments. Tier 1 candidates (UCP3, FBXO32) highlighted.

Fig. 3. Multi-dataset evidence matrix for Tier 1 and Tier 2 candidates. Heatmap: columns = dataset × evidence type; rows = gene symbol; cell colour = magnitude and direction; cell border = significance threshold met. No single dataset drives any candidate's tier assignment.

Fig. 4. Multi-axis pharmacodynamic biomarker panel. (a) Bubble plot of Tier 1/2/3 candidates by biological axis (shape) and tier (colour); bubble size proportional to therapeutic window score. (b) Bar chart of RT_prot_logFC for top 20 candidates. (c) Bar chart of prot_logFC_DMD for the same candidates. CAMK2D = highest RT response; HSPB1 = largest protein deficit.

Fig. 5. Network context: GNN neighborhood and mechanistic convergence. (a) UMAP embedding of GNN node representations with neighborhood genes highlighted. (b) GO enrichment bubble plot for the 623-gene neighborhood. (c) Schematic of mechanistic convergence: ANKRD2 phosphostate as the upstream hub connecting sarcomere integrity, calcium signaling, proteostasis, and metabolic reprogramming axes.

Supplementary Information

Supplementary Table 1. Complete 623-gene GNN neighbourhood with cosine similarity scores and GO annotation.

Supplementary Table 2. Complete tiered biomarker candidate list (Tier 1/2/3) with all evidence variables (n_RNA_sig_down, prot_logFC, RT_prot_logFC, therapeutic window score, biological axis assignment).

Supplementary Table 3. Full dataset inventory: accession numbers, platforms, sample sizes, processing parameters.

Supplementary Table 4. Sensitivity analysis of condition assignment: per-subject MYOM2 and MYOZ2 $\log_2$ intensities, combined mean rank, MYOM2-alone rank, and condition assignments under both approaches. Log_2FC (DMD vs CTR) for all 11 key biomarker candidates under both CTR compositions.

Supplementary Fig. 1. Condition assignment validation: (a) hierarchical clustering dendrogram; (b) sarcomeric marker intensity heatmap by assigned condition; (c) protein matrix UMAP by condition.

Supplementary Fig. 2. MSF database schema and data engineering workflow. SQLite table relationship diagram with row counts and key foreign keys.