Clinical Proteomics of Rhabdomyolysis: Molecular Insights Toward Early Detection

Anna Benedetti

Introduction

Rhabdomyolysis begins by muscle injury resulting from trauma, chemical toxicity, or metabolic stress, leading to the rapid breakdown of skeletal muscle. Its clinical presentation is variable, and complications such as acute kidney injury pose serious risks. Early diagnosis is often delayed due to nonspecific symptoms and limited diagnostic precision. Hence, there is a critical need for sensitive biomarkers to improve early detection and clinical management.

This study aims to characterize the proteomic response to acute muscle injury and its evolution over time, as well as the effects of Growth Hormone treatment. The goal is to identify protein signatures associated with different stages of damage and recovery, correlate them with markers like creatine kinase (CK), and assess their potential to predict outcomes such as dystrophic fiber prevalence. The findings may enable early biomarker discovery, support risk stratification, and inform personalized therapeutic strategies.

Methods

Data preparation

The data available contains 125 samples from 25 separate subjects, collected at 5 time points: before muscle injury, 0 days, 3-5 days, 7 days and 14 days after the injury. 13 subjects received growth hormone treatment, while 12 received a placebo.

As a first step, the missingness was quantified for both proteins and samples. Proteins with missingness >20% and samples with missingness > 30% were removed from the datased. The missingness was confirmed to be not at random but correlated with protein intensity (Spearman $\rho = -0.86$), likely due to instrumentation sensitivity.

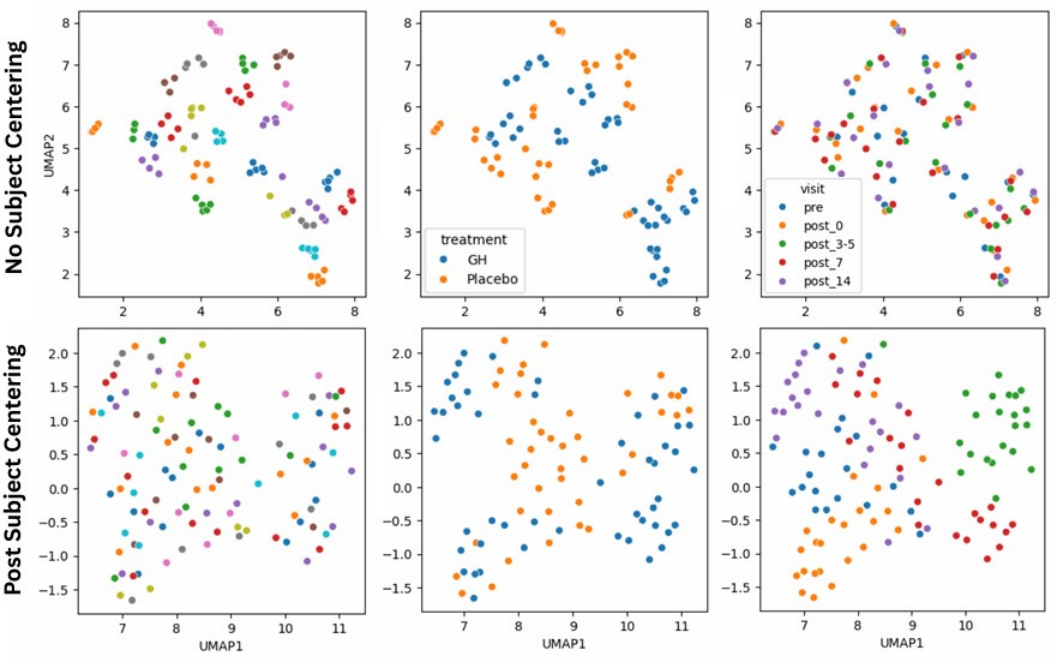
Several imputation strategies from the PIMMS framework were compared, and Collaborative Filtering (CF) was selected. The CF model achieved the lowest validation loss (0.62) against 10 % held-out observed values and preserved the expected left-tail distribution of intensities better than baseline methods (mean, median, random and zero imputation), as well as VAE or DAE methods.

Sample-wise Z-scores of global intensity means were computed to identify outliers. A single sample with $|Z| > 3$, S46 was removed from the dataset.

The data was then Z-score-normalized across samples to remove global loading differences. For exploratory purposes, an additional dataset was prepared by subject-centering all samples (i.e. subtracting each subject's mean profile).

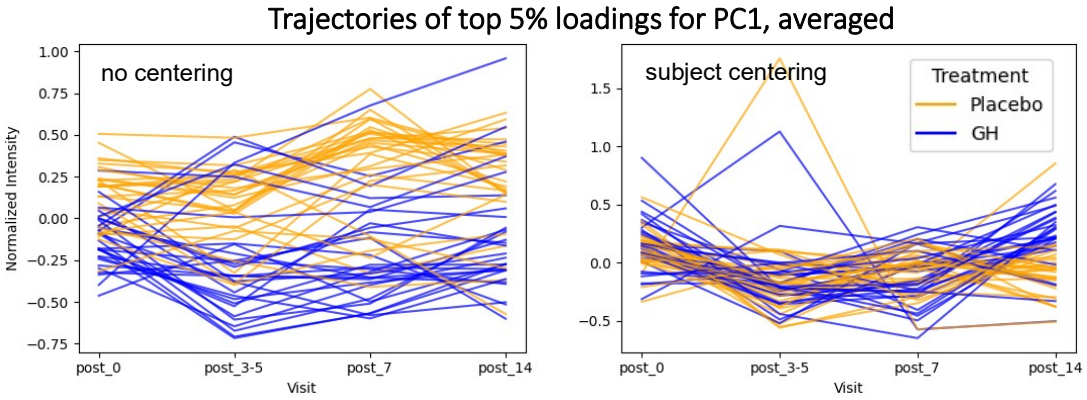
Exploratory data analysis

UMAP (neighbors = 15, minimum distance = 0.1) was carried out on both uncentered and subject-centered data. Before centering, samples mainly clustered by subject. Afterwards, samples arranged primarily by visit, retaining a subtle GH- placebo split.



PCA was also used to analyse how the variance within the data can be attributed to different metavariables before and after subject centering. Before centering, all the first 10 PCs correlate with the subject, implying that the variance is governed by inter-subject variability. When plotting the mean PC value curves over different visits, there seems to be a fairly clear separation between GH and placebo subjects.

After subject centering, the visit variable arises as correlated to 8 out of the first 10 PCs, and subject does not show any significant correlation. However, subject-centering also removes a great part of the variance caused by treatment, as shown by the fact that none of the first 10 PCs were correlated to it, and by the absence of treatment clustering in the PC curves by visit. These findings suggest that in order to study any correlation with treatment it is imperative to use the non centered data.



On the other hand, the number of PCs correlated with creatine kinase (CK) only changed from 4 to 3 upon subject centering, while the number of PCs correlated to dystrophy remained 1 but raised from PC8 to PC5. This suggest that in both the uncentered and centered data, these variables are responsible for a much lower portion of the variance, but that they are less susceptible to subject centering than the other metavariables.

Mixed- effects linear modelling

Mixed-linear models were chosen because each participant contributes repeated plasma samples; a subject-specific random intercept (b_i) captures this intra-person correlation while fixed effects test the biology of interest.

Models were fitted for four questions: (1) global association with CK, (2) change across visits, (3) main effect of Growth-Hormone treatment, and (4) the full visit × treatment interaction to reveal when GH alters the trajectory.

Likelihood estimation was performed independently for each protein, and resulting *p*-values were adjusted across the proteome with the Benjamini–Hochberg FDR procedure (α = 0.05) to control for multiple testing.

Predicting long- term outcomes from acute- phase proteomes

To investigate whether any acute phase (day 0) plasma proteins could predict later outcomes, the analysis was restricted to one sample per participant. For each outcome, namely the creatine-kinase concentration on days 3-5 and the percentage of dystrophic fibres on day 7, the data set therefore comprised at most 25 paired observations. In cases like this, traditional linear models are not feasible due to the extreme $p \gg n$ setting, where the number of predictors (proteins, p > 500) greatly exceeds the number of samples (n < 25). In such instances, linear models are prone to overfitting and unstable coefficient estimates, and they can be unreliable without some form of regularization.

Instead, two different univariate screens were carried out for each protein: Spearman correlation (keeping the variables continuous) and Welch T test (stratifying the CK and dystrophy percentage as high or low depending on the median). The resulting p values were then adjusted with Benjamini–Hochberg FDR procedure (α = 0.05) to control for multiple testing.

Furthermore, Lasso regression, a regularized linear model with L1 penalty, was employed to perform feature selection and reduce overfitting in the high-dimensional regime. Leave-one-out cross-validation was used in model training, as it is particularly suitable for small datasets. However, due to the sample size, additional resampling by bootstrapping was necessary to assess the stability of the selected proteins. Models were fitted across 1000 bootstrap iterations, and selection frequencies were recorded to identify proteins consistently retained across resampled datasets, to provide a measure of their robustness.

Since the creatine kinase concentration and the dystrophy percentage were previously observed to be quite robust to subject centering, all steps described in this section were carried out on both the non centered and the subject-centered data.

Enrichment analysis

To contextualize the proteomic findings in terms of biological function, enrichment analyses were performed using the g:Profiler tool. The analysis was restricted to *Homo sapiens* and included gene ontology terms for Biological Process (BP), Molecular Function (MF), and Cellular Component (CC), as well as KEGG and Reactome (REAC) pathways.

Enrichment was conducted on subsets of proteins identified through the mixed linear models described in the previous sections, namely the model based on CK and the model based on visit and treatment. These were then stratified by the sign of their association (positive or negative correlation) in order to identify biological pathways upregulated and downregulated in relation to muscle damage severity, treatment and recovery. Only terms with FDR below 0.05 were retained for interpretation.

Results

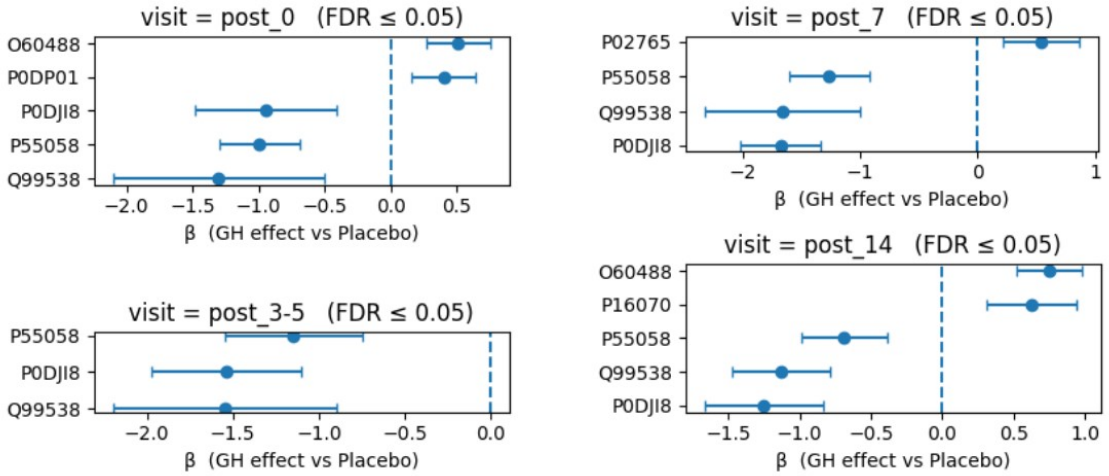
Proteins Exhibiting Treatment-by-Visit Interaction

Linear mixed-effects modelling incorporating a treatment × visit interaction showed that throughout the entire 14 day observation period, three acute-phase proteins, serum amyloid A1 (SAA1, Uniprot P0DJI8), legumain (LGMN, Uniprot Q99538) and phospholipid-transfer protein (PLTP, Uniprot P55058), remained consistently lower in GH-treated participants than in placebo controls. Collectively, this triad ordinarily amplifies cytokine signalling, extracellular-matrix digestion and high density lipoprotein remodelling. Their sustained reduction indicates that GH markedly blunts the classical hepatic acute-phase cascade and restrains macrophage-driven proteolysis, thereby limiting collateral tissue damage and preserving vascular lipid homeostasis.

Against this anti-inflammatory backdrop, GH induced time-locked surges of discrete regenerative mediators. Immediately after injury, plasma showed a transient rise in acyl-CoA synthetase long-chain 4 (ACSL4, Uniprot O60488) and immunoglobulin heavy-chain fragments (IGHV1-8, Uniprot P0DP01), signifying early metabolic priming for fatty acid utilisation, and a rapid humoral immune response that expedites debris clearance. The transition at day 7 was distinguished by a peak in platelet factor 4 (PF4, Uniprot P02776), reflecting heightened platelet activation that stabilises the fibrin scaffold, directs monocyte trafficking and organises the provisional extracellular matrix. By day 14, GH again elevated ACSL4, but also increased the adhesion receptor CD44 (Uniprot P16070), a marker of activated satellite cells and matrix-interacting leukocytes. The late appearance of these proteins denotes intensified membrane biosynthesis and progenitor-cell migration that drive myofibre maturation and architectural remodelling.

This global pattern portrays GH as a temporal coordinator: early suppression of catabolic inflammatory drivers is coupled to punctual activation of metabolic, immunological and structural programmes that hasten the shift from damage control to constructive repair. Clinically, the proteomic signature suggests that GH therapy lowers the risk of systemic complications linked to excessive inflammation while shortening convalescence by promoting efficient haemostasis and satellite cell mediated regeneration.

Differentially Expressed Proteins at Each Phase



Proteins Associated with the Creatine Kinase Concentration

The mixed-effects model with CK as a continuous fixed effect identified 164 proteins whose circulating abundance scaled with muscle-damage severity (FDR < 0.05), 42 of which had a positive correlation, while the remaining 122 had a negative correlation.

Enrichment of the positively correlated proteins was mostly dominated by innate inflammatory mechanisms, such as complement activation, acute-phase response and platelet degranulation. Terms describing extracellular vesicles and leaked contractile structures (sarcomere, myofibril), as well as terms regarding glycolytic and antioxidant categories, were also present. This profile reflects active fibre rupture, release of muscle constituents and mobilisation of innate immune and hemostatic response proportional to injury severity.

Proteins inversely related to CK were characterised by immunoglobulin complex and adaptive immune terms, extensive lipoprotein-particle and lipid-transport categories, and multiple protease-inhibitor activities. This pattern suggests that severe damage is accompanied by a relative depletion of circulating antibody carriers, depression of lipid-transport machinery and loss of endogenous antiproteases, possibly because these resources are consumed or displaced during the inflammatory surge.

Biomarkers of Long Term Muscle Damage

Most of the methods employed to identify potential acute biomarkers for long term muscle damage failed in their aim, likely due to the modest sample size. No method was successful in identifying a predictor for day 7 muscle dystrophy percentage.

In subject-centred data a Lasso model initially retained 14 candidate proteins for the prediction of day 3-5 CK, but none survived bootstrap resampling, underscoring model instability in the $p \gg n$ regime. Univariate screening was largely concordant with this finding, as only one single marker, IGLV2-18, reached significance for later CK (FDR = 0.001, ρ = 0.85).

IGLV2-18 encodes a λ-chain variable domain that participates in antigen recognition within humoral immunity. Elevated circulating immunoglobulin light chains have been linked to complement activation and muscle related autoimmune response. Complement activity is a recognised driver of secondary fibre damage and CK release in rhabdomyolysis and strenuous exercise. Accordingly, this finding suggests that participants mounting an high early humoral response, indexed by high IGLV2-18, subsequently experienced a larger complement-mediated CK surge.
While this single-marker result requires validation in larger cohorts, it implicates early systemic antibody activity, and by extension classical-pathway complement activation, as a potential bridge between initial injury and extended muscle damage.

Conclusions

Mixed linear models found that SAA1, LGMN and PLTP were consistently lower in GH treated subjects than in placebo controls at every sampled time-point, indicating that GH suppressed the acute-phase inflammatory response. GH therapy also produced phase-specific increases: ACSL4 and IGHV1-8 immediately post-injury, PF4 at day 7, and renewed ACSL4 together with CD44 at day 14, suggesting temporally coordinated activation of metabolic, haemostatic and regenerative pathways.
Modelling also showed that creatine kinase rose in concert with complement, platelet and ECM components, confirming that innate immunity and matrix turnover track muscle damage severity. A single early biomarker, IGLV2-18, strongly predicts the later CK peak, implicating brisk humoral activation as a driver of secondary damage.

Statistical power was constrained by the small cohort in analyses using a single data point per subject, leading to unstable multivariate models and high FDR values. Moreover, plasma measurements may not fully mirror intramuscular processes, which in turn could bias enrichment analysis towards blood related terms.
Replication in larger cohorts will be required to validate IGLV2-18 as a biomarker and refine multi protein signatures. To enhance statistical power without imposing the five visit burden, future studies could sample each participant only twice, once acutely and once during the late remodelling phase, thereby facilitating the enrolment of a larger cohort and generating one high-quality data point per subject.

References

1. Webel H, Niu L, Nielsen AB, Locard-Paulet M, Mann M, Jensen LJ, et al. *Imputation of label-free quantitative mass spectrometry-based proteomics data using self-supervised deep learning.* Nat Commun. 2024;15:5405. DOI: 10.1038/s41467-024-48711-5
2. Lazar C, Gatto L, Ferro M, Bruley C, Burger T. *Accounting for the multiple natures of missing values in label-free quantitative proteomics data sets to compare imputation strategies.* J Proteome Res. 2016;15(4):1116–1125. DOI: 10.1021/acs.jproteome.5b00981
3. Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, et al. *g\:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update).* Nucleic Acids Res. 2019;47(W1)\:W191–W198. DOI: 10.1093/nar/gkz369
4. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.* Proc Natl Acad Sci U S A. 2005;102(43):15545–15550. DOI: 10.1073/pnas.0506580102
5. Tibshirani R. *Regression shrinkage and selection via the lasso.* J R Stat Soc Series B. 1996;58(1):267–288. DOI: 10.1111/j.2517-6161.1996.tb02080.x
6. Bosch X, Poch E, Grau JM. *Rhabdomyolysis and acute kidney injury.* N Engl J Med. 2009;361(1):62–72. DOI: 10.1056/NEJMra0801327
7. Nance JR, Mammen AL. *Diagnostic evaluation of rhabdomyolysis.* Muscle Nerve. 2015;51(6):793–810. DOI: 10.1002/mus.24606
8. Frenette J, Cai B, Tidball JG. *Complement activation promotes muscle inflammation during modified muscle use.* Am J Pathol. 2000;156(6):2103–2110. DOI: 10.1016/S0002-9440(10)65081-X
9. Zhang C, Wang C, Li Y, Miwa T, Cui W, Song W-C, et al. *Complement C3a signaling facilitates skeletal muscle regeneration by regulating monocyte function and trafficking.* Nat Commun. 2017;8:2078. DOI: 10.1038/s41467-017-01526-z
10. Doessing S, Heinemeier KM, Holm L, Mackey AL, Schjerling P, Rennie MJ, et al. *Growth hormone stimulates the collagen synthesis in human tendon and skeletal muscle without affecting myofibrillar protein synthesis.* J Physiol. 2010;588(Pt 2):341–351. DOI: 10.1113/jphysiol.2009.179325
11. Tonkin J, Temmerman L, Sampson RD, Gallego-Colon E, Barberi L, Bilbao D, et al. *Monocyte/Macrophage-derived IGF-1 orchestrates murine skeletal muscle regeneration and modulates autocrine polarization.* Mol Ther. 2015;23(7):1189–1200. DOI: 10.1038/mt.2015.66
12. Hattori N. *Expression, regulation and biological actions of growth hormone and ghrelin in the immune system.* Growth Horm IGF Res. 2009;19(3):187–197. DOI: 10.1016/j.ghir.2008.12.001
13. Hernández B, Parnell AC, Pennington SR. *Why have so few proteomic biomarkers "survived" validation? (Sample size and independent validation considerations).* Proteomics. 2014;14(13–14):1587–1592. DOI: 10.1002/pmic.201300377