cardiogenetics

MDPI

Article

Report on the Post-Translational Modifications (PTMs) Prediction in Hypertrophic Cardiomyopathy-Associated Proteins MYH7, MYBPC3, TNNT2, and TNNI3, and Five Unknown PTMs in MYH7 (K129, K1451) and MYBPC3 (K14, R44, T705)

Natasha Trajkovska [1], Lenche Jovova [2] and Done Stojanov [3,*]

[1] School of Informatics, Communications and Media, University of Applied Sciences Upper Austria, 4232 Hagenberg, Austria; natasha.trajkovska@fh-hagenberg.at
[2] Faculty for Computer Science and Engineering, University Ss Cyril and Methodius, 1000 Skopje, North Macedonia; lenchejovova@gmail.com
[3] Faculty of Computer Science, Goce Delcev University, 2000 Stip, North Macedonia
* Correspondence: done.stojanov@ugd.edu.mk

Abstract

In this study, we have performed computational PTM analysis on a panel of hypertrophic cardiomyopathy (HCM)-associated proteins: MYH7, MYBPC3, TNNT2, and TNNI3. We aimed to benchmark the prediction of PTM sites of three ML-based tools: MusiteDeep, PTMGPT2, and SiteTack, using PhosphoSitePlus as a reference for true positives. Notably, because the highest precision tool varied by protein and PTM type, our results indicate there is no single best tool for PTM prediction. Specifically, for HCM-associated proteins, MusiteDeep had the highest precision for MYBPC3 and MYH7; PTMGPT2 was best for TNNI3, and SiteTack for TNNT2. Examining PTM type and phosphorylation in particular, MusiteDeep had the highest precision, followed by PTMGPT2 and SiteTack. However, MusiteDeep did not identify acetylation sites, where PTMGPT2 outperformed SiteTack. Beyond these benchmarking results, we also report on five high-priority candidates for experimental validation in two HCM-associated proteins: MYH7 (K1451 acetylation, K129 methylation) and MYBPC3 (T705 phosphorylation, K14 acetylation, R44 methylation).

Keywords: post-translational modifications; PTM; hypertrophic cardiomyopathy; bioinformatics pipeline; tool benchmarking; phosphorylation; acetylation; methylation; ubiquitination; algorithm comparison; machine learning

Check for updates

Academic Editor: Giuseppe Limongelli

Received: 30 December 2025
Revised: 13 March 2026
Accepted: 26 March 2026
Published: 2 April 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Post-translational modifications (PTMs) are changes in protein properties after they are synthesized, affecting both structure and function. These modifications add new properties that go beyond the initial role set by amino acids' basic characteristics, greatly influencing various cellular processes, such as modulations in protein and molecular interactions, the regulation of gene expression, localization, and signaling [1] (pp. 255–261).

There are two categories of PTMs. The first category of PTM includes changes resulting from the covalent addition of a modifying group, such as methyl, phosphoryl, acetyl, or glycosyl. It is very important to note that these modifications are reversible. The second category refers to alterations resulting from proteolytic cleavage (proteolysis), in which peptide bonds are broken to yield smaller peptides or amino acids [2,3]. According to Bradley [4], advances in mass spectrometry have contributed to the discovery of over

600 different PTM classes and (Ramazi & Zahiri, 2021) [2] highlighted the ten most studied PTMs, such as phosphorylation, acetylation, ubiquitination, methylation, glycosylation, SUMOylation, palmitoylation, myristylation, prenylation, and sulfation. Phosphorylation, methylation, acetylation, and ubiquitination are the focus of this study. These PTMs are the most frequently occurring PTMs in inherited human diseases.

Phosphorylation, the process of adding a phosphate group (PO_4^{3-}) from ATP, is one of the most studied types of PTM. Target residues for phosphorylation include S (serine), T (threonine), and Y (tyrosine) [5]. Phosphorylation controls cellular functions, such as signal transduction, metabolism, and cell division [6]. In the context of cardiomyopathy, phosphorylation has an impact on calcium sensitivity in the heart tissue that leads to slow muscle contraction and eventually heart failure [7].

Methylation, like phosphorylation, is also a reversible process that is characterized by the addition of a methyl group (-CH3) to target amino acids. The process is facilitated by methyltransferases, and the donor of the methyl group is S-Adenosyl Methionine (SAM) [2,3]. Candidate residues for methylation are K (lysine) and R (arginine). Histones are the prime target for methylation. From a biomedical perspective, methylation affects chromatin organization and gene expression, with a primary focus on gene regulation. Genetic changes associated with cardiomyopathy may lead to alterations in heart tissue structure and malfunction [8].

Acetylation is a process of adding an acetyl group ($COCH_3$) to the target amino acid, which is usually lysine (K). This process is catalyzed by lysine acetyltransferases (KATs), and acetyl-CoA serves as the donor. Similar to methylation, acetylation targets histones and affects gene expression and chromatin structure [9]. Modified acetylation has been associated with cardiovascular diseases, such as cardiomyopathy. The presence of acetylation on proteins involved in mitochondrial function may alter energy metabolism, resulting in an energy deficit in heart cells. Acetylation also has an impact on the transcription factors responsible for cardiac hypertrophy and fibrosis [10].

Ubiquitination is a PTM where a protein called ubiquitin, containing 76 amino acids, is attached to a target protein. This is a cascade process aided by different enzyme types, such as E1 or ubiquitin-activating enzyme, E2 or ubiquitin-conjugating enzyme, and E3 or ubiquitin ligases [11]. Ubiquitination occurs mostly on K (lysine) and can cause proteasomal degradation, DNA repair, endocytosis, and signal transduction [12]. The result of these processes in the context of cardiomyopathy is that they contribute to heart inflammation and fibrosis, which affects heart muscle remodeling and failure [13].

The biological importance of PTMs has led to the creation of computer- and AI-driven approaches for their identification. Several ML tools, including MusiteDeep, PTMGPT2, and SiteTack, have been designed for the computational detection of PTMs. Nonetheless, it is crucial to note that the results of these models do not always align with experimental evidence.

HCM is a genetic cardiac disorder characterized by unexplained thickening of the myocardium, most often the interventricular septum, in the absence of other causes, such as hypertension or valvular disease [14]. The pathology is characterized by several changes in the heart tissue, including myocyte hypertrophy and disarray (enlargement and disorganization of heart muscle cells), interstitial fibrosis (scarring between cells), and small-vessel disease (impairment of small blood vessels) [15]. In some cases, left ventricular outflow tract (LVOT) obstruction is also present, leading to obstruction of the heart's main blood flow [16]. Some cases are asymptomatic, others experience dyspnea, chest pain, or fainting. Potential complications are atrial fibrillation (raising the risk of stroke), ventricular arrhythmias, and ultimately, sudden cardiac death [17]. Although HCM is primarily caused by

mutations in sarcomeric proteins, PTMs can strongly modify disease severity, progression, and phenotype, thereby representing an important area of investigation [18].

In this paper, we aimed to benchmark the predictive PTM performance of three ML-based applications: MusiteDeep, PTMGPT2, and SiteTack for PTM detection, specifically for phosphorylation, methylation, acetylation, and ubiquitination on a small panel of sarcomeric proteins, commonly associated with HCM [19]. The analysis of PTM distribution in the human body, based on PTMD 2.0 [20], follows our PTM types selection: phosphorylation is the most frequent modification, followed by ubiquitination, acetylation, and methylation (Figure 1).

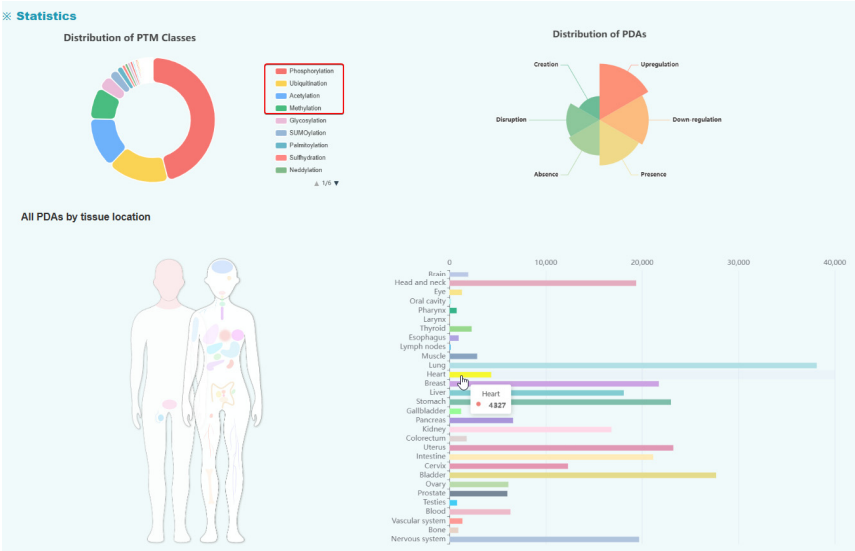


Figure 1. Distribution of PTM classes according to PTMD 2.0, with emphasis on the four most frequent classes.

The experiments utilized a dataset comprising protein sequences from the *MYH7*, *MYBPC3*, *TNNT2*, and *TNNI3* genes, which encode key cardiac sarcomere proteins. MYH7 encodes the β-myosin heavy chain, a major motor protein of the cardiac thick filament vital for cardiac force generation. *MYBPC3* encodes cardiac myosin-binding protein C, which supports sarcomere structure and regulates cardiac contraction and relaxation. *TNNT2* (cardiac troponin T) and *TNNI3* (cardiac troponin I) are components of the troponin complex on the thin filament, which modulates calcium-dependent contraction [21]. Additionally, this study examined whether any novel PTM sites are consistently predicted by three machine learning (ML) tools but not yet experimentally confirmed in PhosphoSitePlus. We identified five high-priority PTM candidates suitable for targeted experimental validation: MYH7 (K1451 acetylation, K129 methylation) and MYBPC3 (T705 phosphorylation, K14 acetylation, R44 methylation).

2. Materials and Methods

In this study, we aim to evaluate the performance of three ML applications: MusiteDeep, developed by researchers at University of Missouri-Columbia, USA and Jilin University, China (https://www.musite.net/, accessed on 5 September 2025), PTMGPT2, hosted via the National Supercomputing Center for Life Sciences, Jeonbuk National University, South Korea (https://nsclbio.jbnu.ac.kr/tools/ptmgpt2/, accessed on 5 September 2025), and SiteTack, developed by researchers at Massachusetts Institute of Technology and Broad Institute of MIT and Harvard (https://sitetack.net/, accessed 5 September 2025) through a case study of four proteins associated with HCM: MYH7, MYBPC3, TNNT2, and TNNI3.

Additionally, we have explored their ability to discover undiscovered PTM sites that could be designated as high-priority candidates for further experimental validation. In Table 1, we provide details on protein full names, NCBI RefSeq ID, length, gene-coding symbol and chromosomal location. We aimed to benchmark the PTM predictive performance of MusiteDeep, PTMGPT2, and SiteTack across four PTM types: phosphorylation, acetylation, ubiquitination, and methylation. MusiteDeep and SiteTack are trained on large collections of PTM-annotated protein sequences (primarily from Uniport/Swiss-Prot and previously curated PTM datasets) rather than on specific protein families, covering multiple modification types [22,23]. PTMGPT2 reports a defined large-scale training dataset of roughly 388,000 annotated PTM instances used to fine-tune its language model [24]. We aim to provide comprehensive predictive performance evaluations for each sarcomeric HCM-associated protein and for each PTM type.

Table 1. Details of the selected HCM-associated proteins retrieved from NCBI and HGNC.

Protein	NCBI RefSeq ID	Length (aa)	Gene Symbol	Chromosomal Location
MYH7 (Cardiac β-myosin heavy chain)	NP_000248.2	1935	MYH7	14q11.2
MYBPC3 (Myosin-binding protein C, cardiac)	NP_000247.2	1274	MYBPC3	11p11.2
TNNT2 (Troponin T, cardiac type)	NP_001263274.1	298	TNNT2	1q32.1
TNNI3 (Troponin I, cardiac type)	NP_000354.4	210	TNNI3	19q13.4

We used PhosphoSitePlus v6.8.1 (https://www.phosphosite.org/) as a benchmarking reference, which is a manually curated, interactive database that focuses on experimentally observed PTMs [25]. PhosphoSitePlus is a widely used PTM annotation resource that covers a vast number of PTM sites across various types. It includes hundreds of thousands of unique, non-redundant modification sites, curated from more than 22,000 articles and numerous mass spectrometry datasets [26] Figure 2 shows an example visual output of the PhosphoSitePlus tool for the MYH7 protein.

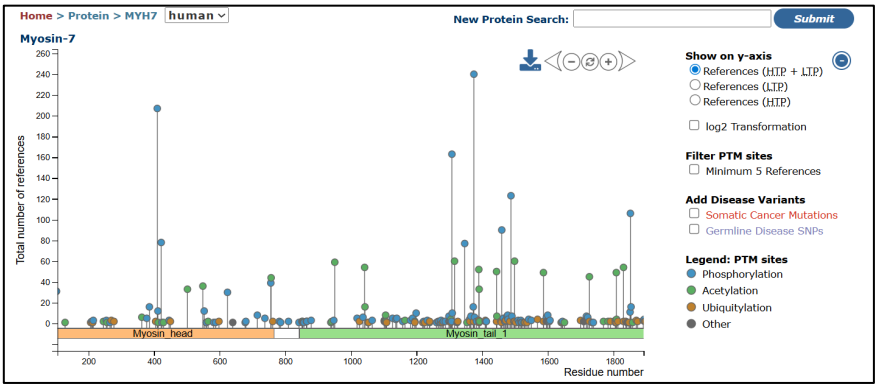


Figure 2. A sample visual output provided by PhosphoSitePlus.

The selection of the PTM prediction tools included in the benchmarking process was based on three criteria:

(1) They can provide coverage of multiple PTMs at once.
(2) They can be accessed directly as a web application.
(3) They employ contemporary ML methodologies.

This approach is intended to ensure that selected tools are both efficient and easy to use. Based on criteria 1–3, we have chosen: MusiteDeep [22], PTMGPT2 [24] and SiteTack [23]. MusiteDeep utilizes deep learning, specifically convolutional neural networks (CNNs) with attention mechanisms, to predict PTM sites in proteins [22]. PTMGPT2 is based on the GPT2 architecture and employs a transformer-based approach, fine-tuned with prompt-based learning to predict PTM sites [24], while SiteTack incorporates deep learning models that utilize known PTM sites as part of the input encoding, enhancing prediction accuracy [23]. Table 2 provides an overview of the selected tools.

Table 2. Overview of the selected ML tools for PTM prediction.

Tool	URL	ML Method
MusiteDeep	https://www.musite.net/	Deep learning: CNNs with attention mechanisms
PTMGPT2	https://nsclbio.jbnu.ac.kr/tools/ptmgpt2/ (accessed on 5 September 2025)	Based on the GPT-2 architecture: transformer-based approach, fine-tuned with prompt-based learning
SiteTack	https://sitetack.net/ (accessed on 5 September 2025)	Deep learning models that incorporate known PTM sites as part of the input

PhosphoSitePlus records HTPs (High Throughput Papers), representing modification sites reported solely through proteomic discovery mass spectrometry, and LTPs (Low Throughput Papers), indicating sites identified by other methods. To enhance confidence and reliability, we considered PTM sites from PhosphoSitePlus that appeared in more than one HTP study (HTP > 1) as true positives for evaluating our reference set.

We used a probability threshold of 0.6 to predict positives in both the MusiteDeep and SiteTack models, aligning with the models' confidence levels to facilitate comparison with the PhosphoSitePlus reference sites (HTP > 1). The PTMGPT2 tool employs a generative language model, which does not produce explicit prediction scores; thus, we relied on its default output without thresholding. The ML tools were compared using the following metrics: precision, recall, and F1-score. Precision indicates the proportion of correct positives among all predicted positives, recall shows the percentage of actual positives correctly identified, and F1-score balances both. These metrics offer a comprehensive performance assessment, especially in datasets with class imbalance.

Our methodological framework consists of several steps, performed for each protein of interest, including: downloading FASTA files from NCBI, obtaining prediction sites from PhosphoSitePlus, as well as from the three ML tools, gathering and preprocessing the data, and at the end calculating the statistics and visualizing the findings (Figure 3). The pipeline has been implemented in Python 3.10, using pandas, seaborn and matplotlib as key libraries.

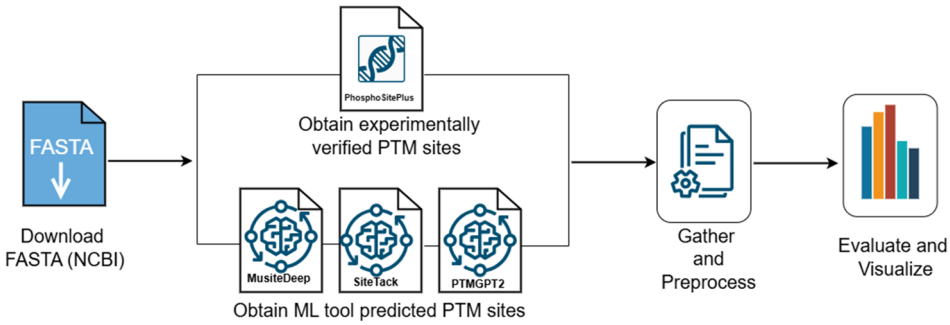


Figure 3. Schematic overview of the applied methodological framework.

3. Results

Four PTM types—phosphorylation, acetylation, ubiquitination, and methylation—and four proteins—MYH7, MYBPC3, TNNT2, and TNNI3—are in the focus of the present research. Apart from the computational analysis on precision, recall, and F1-scores, we have also analyzed the rate of experimentally verified PTM locations, given a different HTP cutoff, as well as the predicted PTM locations with a varying threshold. Finally, we performed an analysis on the false positives.

3.1. PTM Identification with Threshold Adjustment

Figure 4 shows the number of experimentally verified PTM locations for each target protein of interest, for different HTP cutoffs. For the MYH7 protein, in the absence of any HTP filtering, PTM sites were identified for phosphorylation (105 sites), acetylation (42 sites), ubiquitination (2 sites), and methylation (1 site), Figure 4a. When an HTP cutoff greater than 1 was applied, the number of detected sites decreased to 83 for phosphorylation, 34 for acetylation, 1 for ubiquitination, and methylation sites were no longer detected (Figure 4a). With further increases in the HTP cutoff threshold, the number of PTM sites continued to decline, retaining only phosphorylation and methylation sites. In MYBPC3, only acetylation and phosphorylation sites were observed. One acetylation site has been experimentally verified for MYBPC3, independent of the HTP cutoff, whereas the number of experimentally detected phosphorylation sites decreased as the HTP cutoff increased (Figure 4b). A total of 22 phosphorylation sites were validated by at least one HTP, decreasing to 19 sites at HTP > 2 and 15 sites at HTP > 3, and continuing to decline with increasing HTP cutoff, with only 9 phosphorylation sites at HTP > 10 (Figure 4b). For the TNNT2 protein, two experimentally verified PTMs were identified: acetylation and phosphorylation (Figure 4c). As the HTP cutoff increased, only phosphorylation sites remained, with 5 sites detected at HTP > 1 and only a single phosphorylation site at HTP > 2 (Figure 4c). In the TNNI3 protein, 3 types of PTMs were observed: phosphorylation, acetylation, and methylation, corresponding to 17, 6, and 2 sites, respectively (Figure 4d). Applying an HTP cutoff greater than 1 reduced these numbers to 15 phosphorylation, 5 acetylation, and 1 methylation site (Figure 4d). As the HTP threshold increased further, a decline in modification sites was observed, leaving 12 phosphorylation and 3 acetylation sites at HTP > 2, and ultimately only phosphorylation sites at higher cutoffs (Figure 4d).

The results from PhosphoSitePlus, with cutoff HTP > 1, which was selected as the threshold for further analysis to ensure higher confidence in site verification, showed that not every PTM type can be found in all four proteins. Phosphorylation was the most frequent PTM type and was observed in all four proteins (Figure 4). Acetylation is identified in MYH7, MYBPC3, and TNNI3, as shown in Figure 4a,b,d. Ubiquitination is identified only in MYH7 (Figure 4a), while methylation is identified only in TNNI3 (Figure 4d).

MusiteDeep predicted 54 phosphorylation sites, 12 acetylation sites, 32 ubiquitination sites, and 1 methylation site for the MYH7 protein, at a threshold of 0.5. For a threshold of 0.6, the numbers changed to 48, 4, 13, and 1, respectively (Figure 5a). As the threshold increased, only phosphorylation sites (showing a gradual decline in number) and a single methylation site remained (Figure 5a). Notably, the methylation site was not reported for this protein in PhosphoSitePlus (Figure 4a). For MYBPC3 protein, MusiteDeep predicted 28 points of phosphorylation, 14 acetylation, 10 ubiquitination, and 2 methylation for a threshold of 0.5 (Figure 5b). When the PTM score was set to 0.6, the corresponding counts were 20, 8, 2, and 2, and kept decreasing with the increased threshold (Figure 5b). Although all four PTM types were detected for TNNT2 at a 0.5 threshold, increasing the threshold resulted in a reduction to only phosphorylation sites: 10 at 0.6, 6 at 0.7, and 1 at 0.8. For TNNI3, MusiteDeep predicted 17 sites of phosphorylation, 1 point of methylation, and

1 site of ubiquitination for a cutoff score of 0.5 (Figure 5d). For a PTM cutoff score of 0.6 and higher, only phosphorylation sites were predicted (Figure 5d).

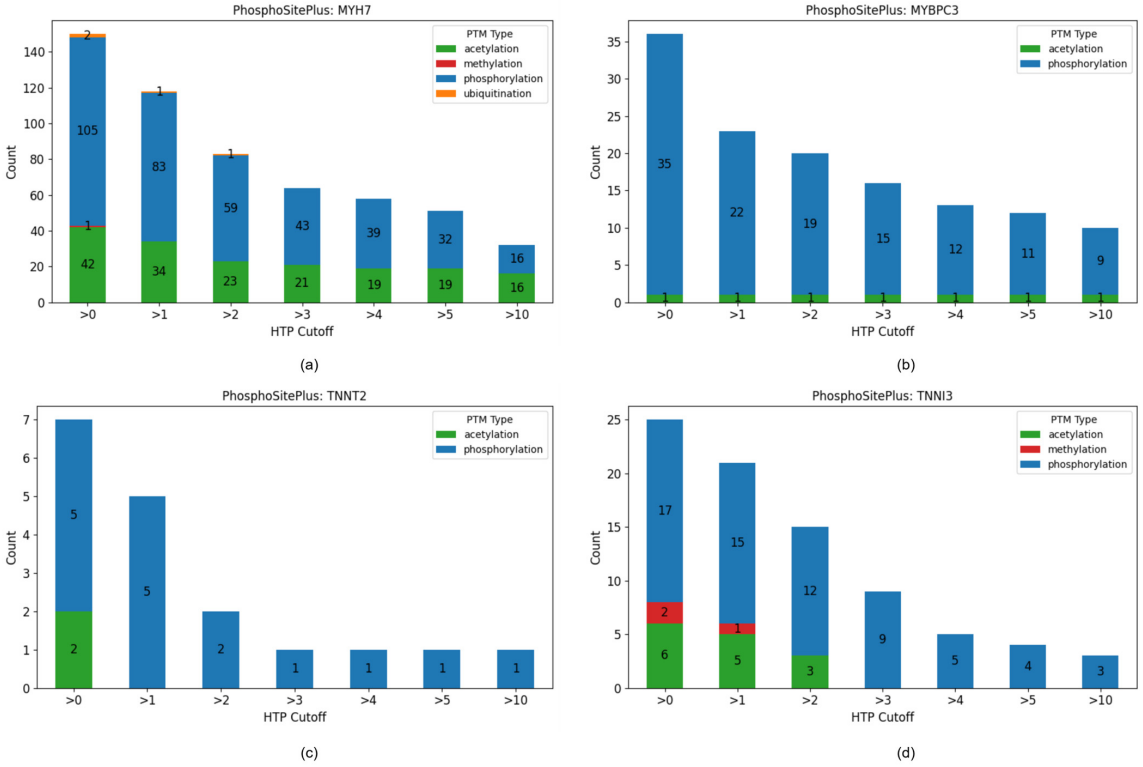


Figure 4. Number of PTM sites reported by PhosphoSitePlus for: (**a**) MYH7; (**b**) MYBPC3, (**c**) TNNT2; (**d**) TNNI3.

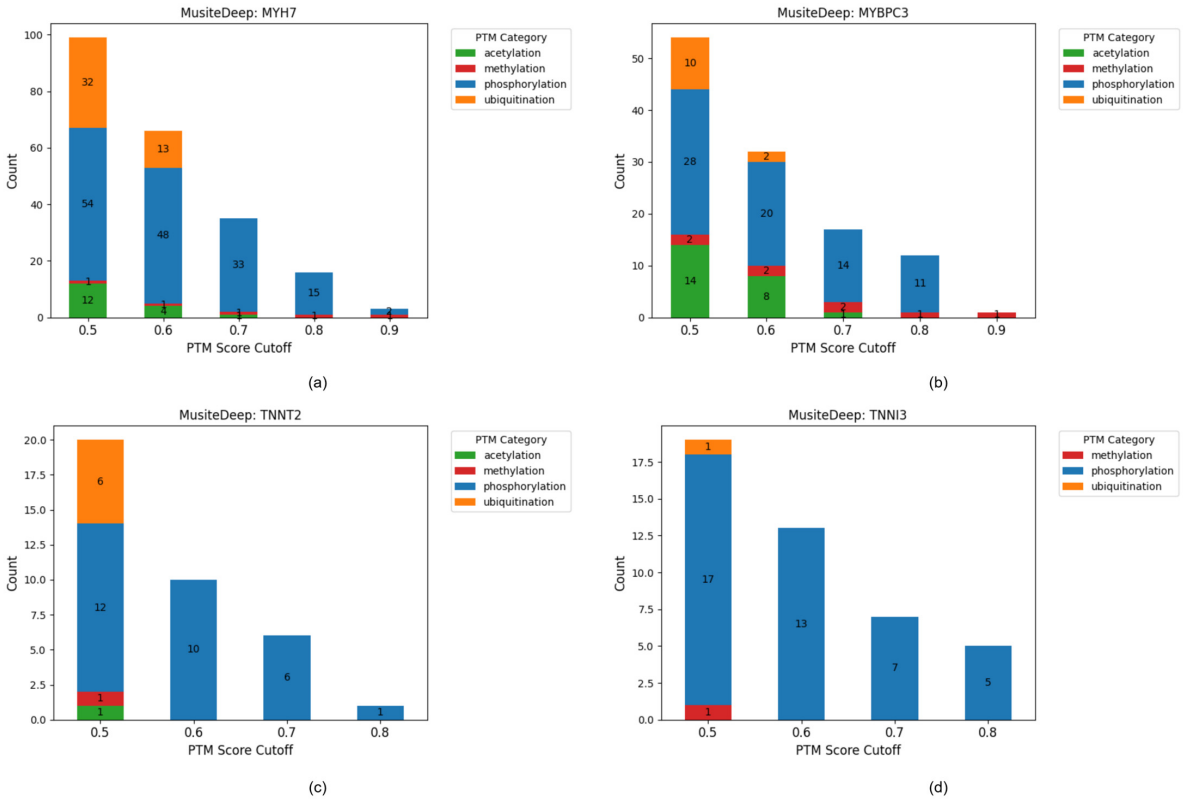


Figure 5. Number of PTM sites predicted by MusiteDeep for: (**a**) MYH7; (**b**) MYBPC3, (**c**) TNNT2; (**d**) TNNI3.

Figure 6 shows the number of PTM sites predicted by SiteTack. For MYH7, phosphorylation, acetylation, ubiquitination, and methylation sites were identified at a probability score cutoff of 0.6, with counts of 78, 40, 21, and 13, respectively (Figure 6a). The number of predicted sites was higher at a score of 0.5 and decreased with increasing probability thresholds. Similar patterns were observed for MYBPC3, TNNT2, and TNNI3. At a probability score of 0.6, MYBPC3 was predicted to have 36 phosphorylation, 24 acetylation, 17 ubiquitination, and 14 methylation sites (Figure 6b). TNNT2 has 12 phosphorylation, 6 acetylation, 2 ubiquitination, and 10 methylation sites (Figure 6c), while TNNI3 was predicted to contain 15 phosphorylation, 6 acetylation, 8 ubiquitination, and 7 methylation sites at the same threshold.

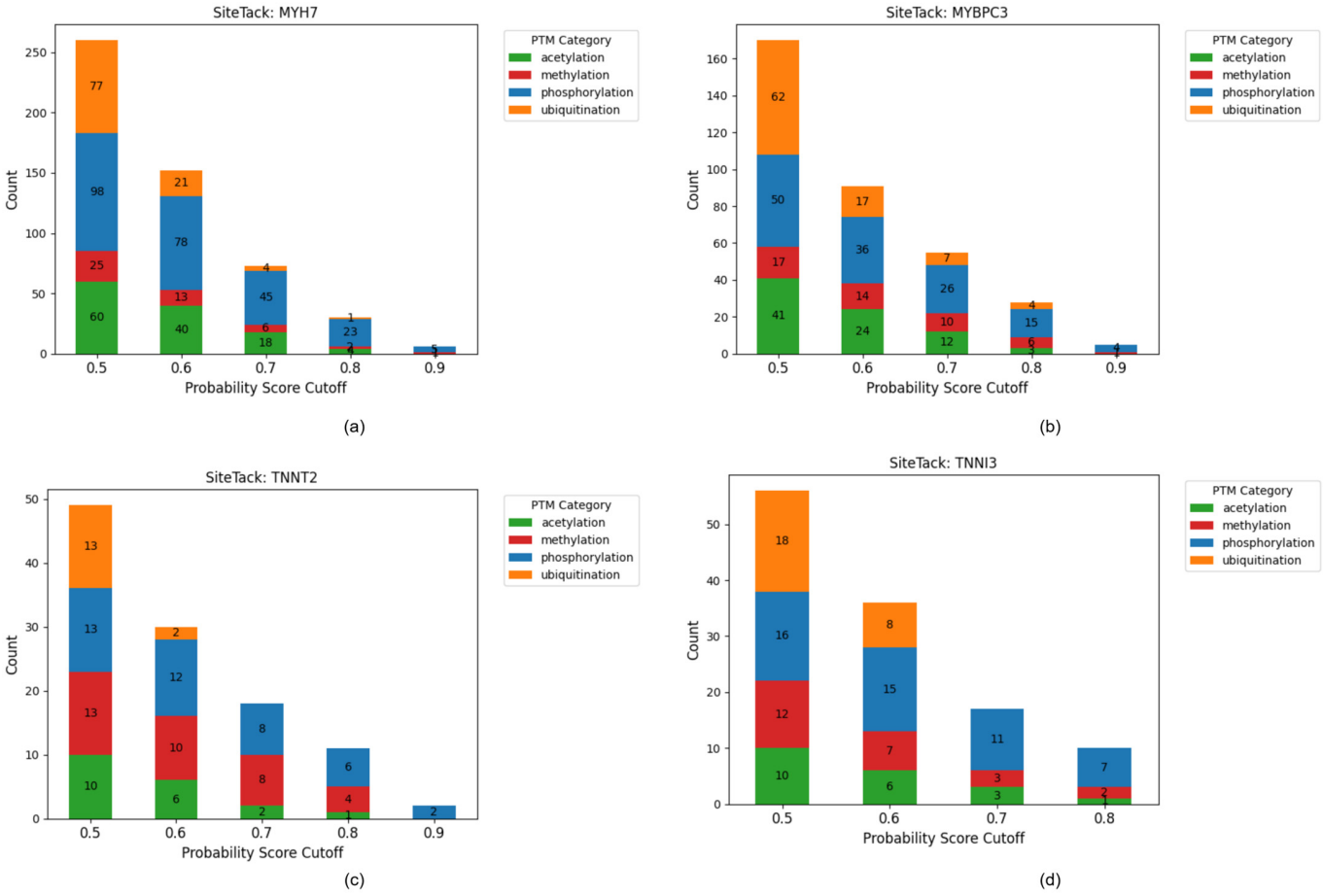


Figure 6. Number of PTM sites predicted by SiteTack for: (**a**) MYH7; (**b**) MYBPC3, (**c**) TNNT2; (**d**) TNNI3.

The previously observed trend is consistent across both experimental and ML approaches: applying more rigorous cutoffs reduces the number of identified PTM sites.

Figure 7 shows the number of PTM sites predicted by PTMGPT2 for the target proteins. For MYH7, 38 phosphorylation sites, 58 acetylation sites, 6 ubiquitination sites, and 4 methylation sites were detected. MYBPC3 was predicted to contain 17 phosphorylation, 3 acetylation, 1 ubiquitination, and 3 methylation sites (Figure 7). No methylation sites were detected for TNNT2 or TNNI3 (Figure 7). For these proteins, PTMGPT2 predicted 5 phosphorylation, 5 acetylation, and 1 ubiquitination point for TNNT2, while TNNI3 had 8 phosphorylation, 2 acetylation, and 1 ubiquitination site.

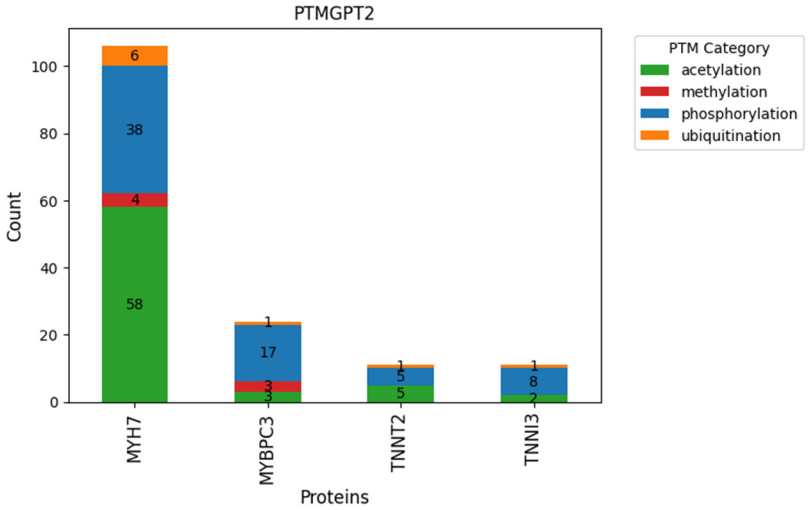


Figure 7. Number of PTM sites predicted by PTMGPT2.

3.2. Evaluation of Predictive Performance

For the benchmarking purpose, we have taken PhosphoSitePlus experimentally verified PTM locations for HTP > 1 as true positives, while the threshold for predicted PTM positives was set to 0.6 in both ML applications, MusiteDeep and SiteTack. For PTMGPT2, we have used the model's default output, as no cutoff score needed to be provided.

We have calculated the precision, recall, and F1-score for each protein, and then we have calculated the average (per-protein) statistics, presented on Figure 8. In terms of phosphorylation, which was the dominant PTM type, all tools showed moderate and statistically not different precision, with MusiteDeep achieving the highest, 0.554, 95% CI (0.137, 0.920), followed by PTMGPT with 0.527, 95% (0.099, 0.926), and SiteTack with 0.511, 95% CI (0.209, 0.806), as shown in Figure 8a. SiteTack demonstrated the highest recall of 0.691 (Figure 8a). Accordingly, SiteTack showed the best balance between precision and recall, achieving an F1-score of 0.565 (Figure 8a). MusiteDeep failed to identify acetylation sites, while PTMGPT2 showed better performance compared to SiteTack (Figure 8b). Average statistics for ubiquitination were relevant only for SiteTack, which demonstrated a very low precision of 0.012 and maximum recall (Figure 8c). The summary statistics for the methylation, represented by only 1 PTM location in TNNI3 protein, according to the experimentally verified results, did not provide any meaningful overview, since all tools failed to detect it (Figure 8d).

Figures 9 and 10 provide a closer insight into the results for phosphorylation and acetylation per protein, and the detailed information is presented in Table 3, which contains the calculated statistics for the relevant proteins and PTM types. As we can see from the results in Figure 9, MYBPC3 and TNNT2 were more problematic for prediction compared to MYH7 and TNNI3. MusiteDeep had the highest precision for MYBPC3 and MYH7, PTMGPT2 was the most precise for TNNI3, while SiteTack was the most precise for TNNT2 (Figure 9a). In general, SiteTack showed the highest recall, reaching 1 for TNNT2 (Figure 9b). Based on the F1-scores for phosphorylation (Figure 9c), MusiteDeep and SiteTack outperformed PTMGPT2. The highest F1-score was observed for TNNI3, 0.786 by MusiteDeep, but for all other proteins, SiteTack's F1-score was higher (Table 3).

Figure 10 confirms that the best results for acetylation were provided by PTMGPT2 with an F1-score of 0.652 for MYH7, and 0.571 for TNNI3, compared with the results from SiteTack, with an F1-score of 0.162 and 0.363, respectively (Table 3).

Ubiquitination was only identified by SiteTack for MYH7, with a precision of 0.048, a recall of 1 and an F1-score of 0.09 (Table 3).

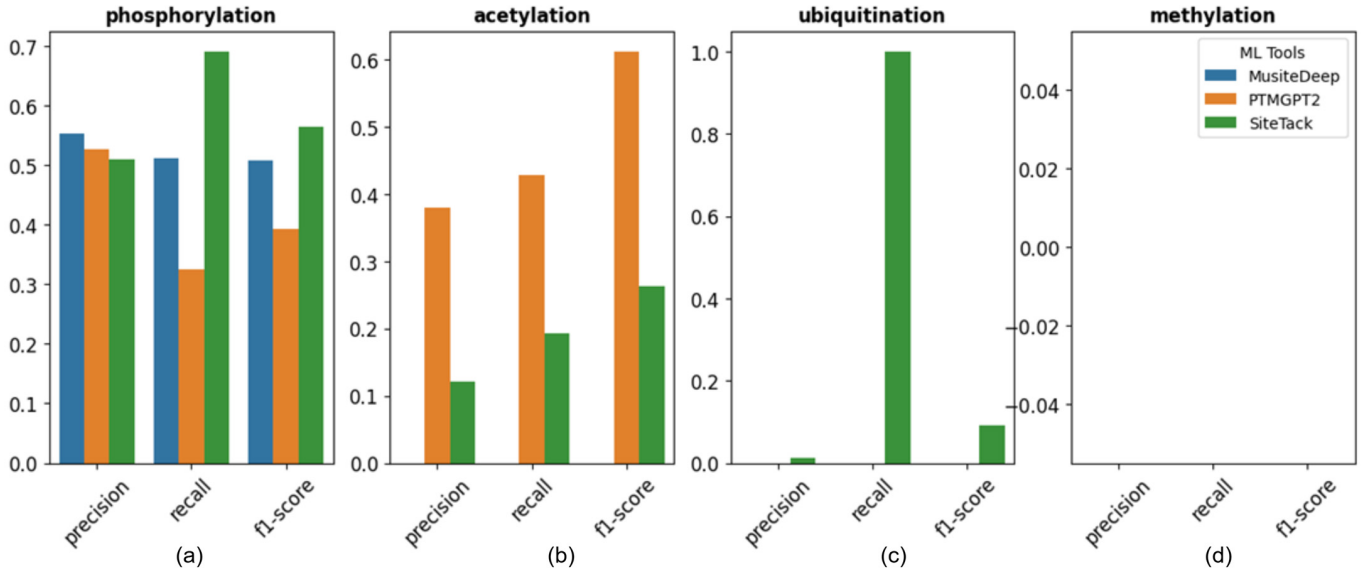


Figure 8. Average precision, recall, and F1-score calculated across four proteins (MYH7, MYBPC3, TNNT2, and TNNI3) for (**a**) phosphorylation, (**b**) acetylation, (**c**) ubiquitination and (**d**) methylation.

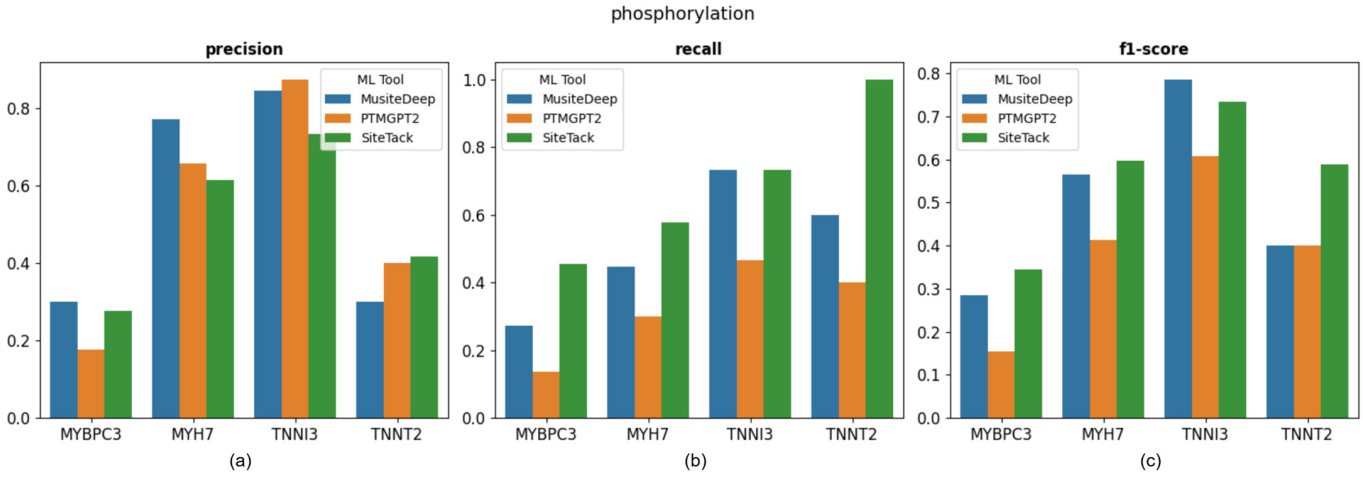


Figure 9. Precision (**a**), recall (**b**), and F1-score (**c**) for phosphorylation across individual proteins.

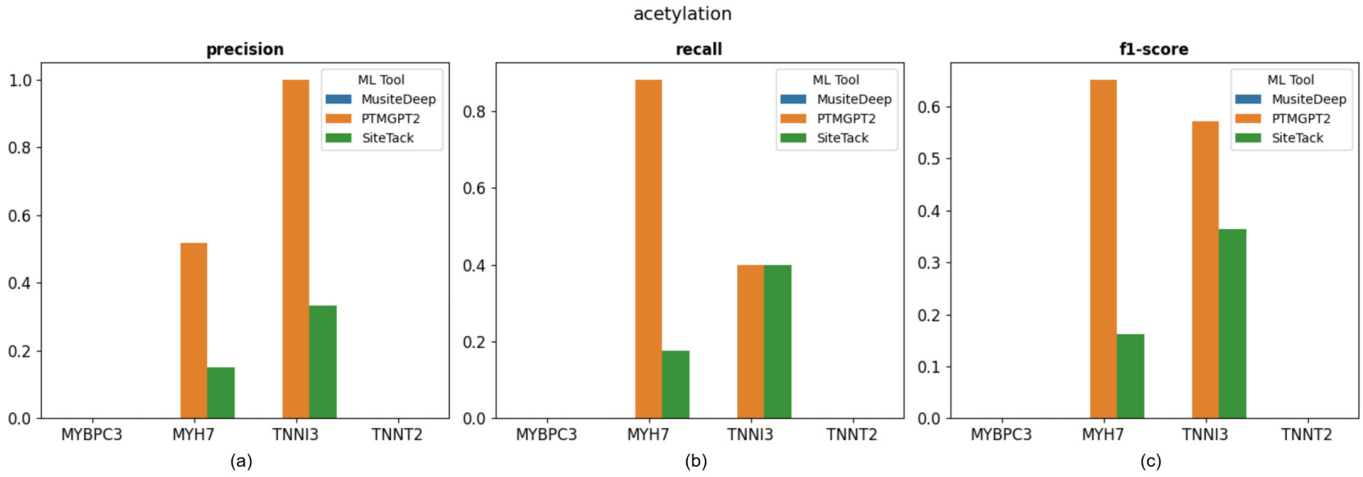


Figure 10. Precision (**a**), recall (**b**), and F1-score (**c**) for acetylation across individual proteins.

Table 3. Summary of precision, recall and F1-score per protein and PTM type.

		Precision			Recall			F1-Score		
	Protein	**Musite Deep**	**PTMGPT2**	**SiteTack**	**Musite Deep**	**PTMGPT2**	**SiteTack**	**Musite Deep**	**PTMGPT2**	**SiteTack**
phosphorylation	MYBPC3	0.300	0.176	0.278	0.273	0.136	0.455	0.286	0.154	0.345
	MYH7	0.771	0.658	0.615	0.446	0.301	0.578	0.565	0.413	0.596
	TNNI3	0.846	0.875	0.733	0.733	0.467	0.733	0.786	0.609	0.733
	TNNT2	0.300	0.400	0.417	0.600	0.400	1.000	0.400	0.400	0.588
acetylation	MYH7	0.000	0.517	0.150	0.000	0.882	0.176	N/A *	0.652	0.162
	TNNI3	N/A *	1.000	0.333	0.000	0.400	0.400	N/A *	0.571	0.364
ubiquitination	MYH7	0.000	0.000	0.048	0.000	0.000	1.000	N/A *	N/A *	0.091

* N/A indicates that the metric is undefined due to the absence of predicted positives or ground truth positives.

Because MusiteDeep and SiteTack output probabilistic scores for PTM predictions, precision–recall (PR) and Receiver Operating Characteristic (ROC) curves for the main PTM types, phosphorylation and acetylation, were computed. Figure 11 presents aggregated PR and ROC curves across the four analyzed proteins (MYH7, MYBPC3, TNNT2 and TNNI3), along with the corresponding AUC–ROC and AUC–PR values for each tool. Aggregated ROC and precision–recall analyses across all proteins revealed moderate predictive performance for phosphorylation sites for both tools. ROC analysis showed AUC–ROC values of 0.752 and 0.757 for MusiteDeep and SiteTack, respectively (Figure 11a,c), indicating comparable discriminative ability. Consistently, precision–recall analysis (Figure 11b,d) showed moderate performance for phosphorylation prediction (AUC–PR = 0.50–0.57), with MusiteDeep showing a slight improvement over SiteTack. In contrast, acetylation prediction performance was weak for both tools. ROC analysis produced AUC–ROC values close to random, 0.497 and 0.452 (Figure 11a,c), and precision–recall analysis similarly indicated poor predictive performance of AUC–PR $\approx$ 0.11 (Figure 11b,d). These results suggest that, while both predictors can moderately discriminate phosphorylation sites, they have limited ability to distinguish acetylated residues from non-modified sites in this dataset, likely reflecting the extreme class imbalance and the limited availability of experimentally validated acetylation sites.

We have additionally computed F1-optimal thresholds based on aggregated predictions across the analyzed proteins. In Table 4, we summarize the AUC–ROC and AUC–PR values, as well as the best F1 thresholds, computed for both PTM types (phosphorylation and acetylation) and for both tools (MusiteDeep and SiteTack).

Table 4. Aggregated performance metrics and F1-optimal thresholds for MusiteDeep and SiteTack.

Tool	**PTM Type**	**AUC–ROC**	**AUC–PR**	**Best F1 Threshold**
MusiteDeep	phosphorylation	0.752	0.573	0.329
MusiteDeep	acetylation	0.497	0.117	0.201
SiteTack	phosphorylation	0.757	0.501	0.372
SiteTack	acetylation	0.451	0.113	0.167

Although the threshold that maximizes the F1-score is lower (around 0.3), we have selected a classification threshold of 0.6 to prioritize prediction confidence. In the context of PTM site prediction, false positives can lead to unnecessary experimental validation, which is costly and time-consuming. Using a higher threshold increases precision at the expense of recall. This more conservative setting is further justified by the definition of the ground truth used in this study, which required true positive PTM sites to be supported by more than one independent experiment. Because this criterion favors highly reliable annotations and may exclude real but less frequently observed modifications, we opted for a stricter

threshold to focus on the most confident predictions and produce a smaller, more reliable set of candidate PTM sites for downstream validation.

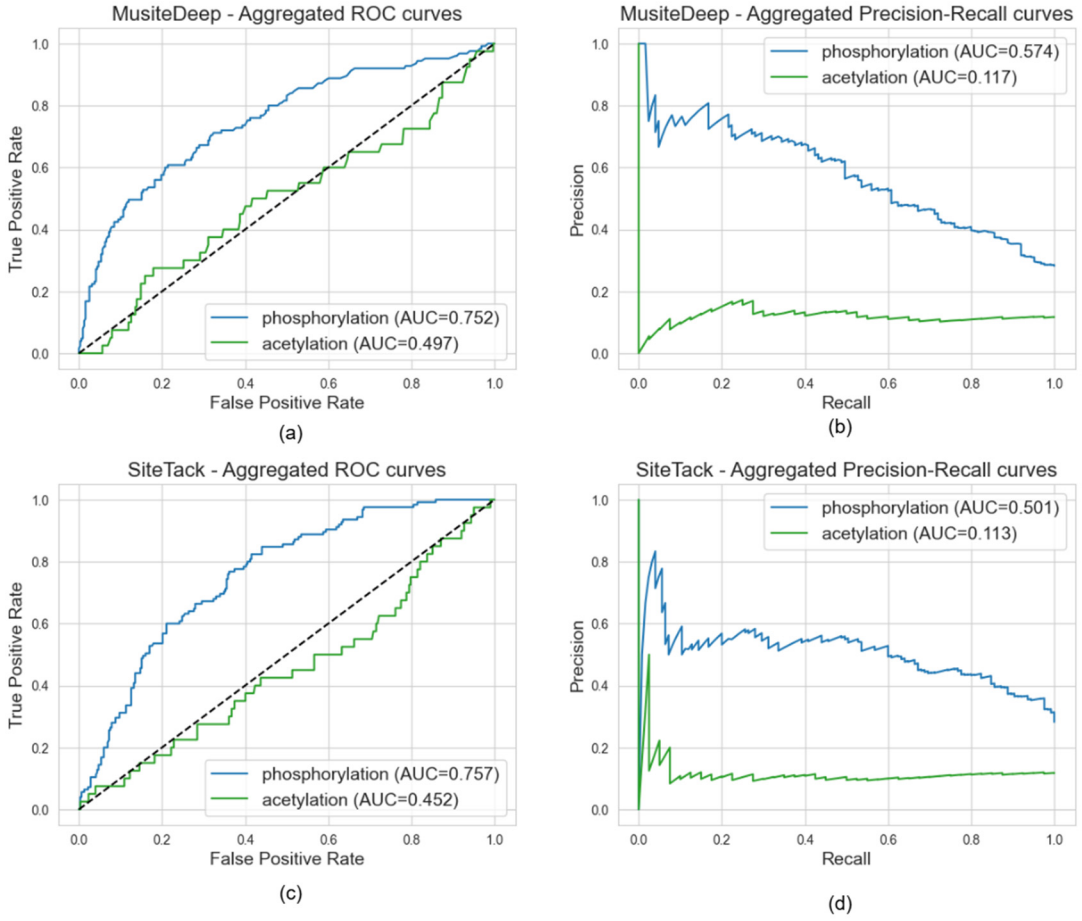


Figure 11. Aggregated ROC and precision–recall curves for phosphorylation and acetylation with the corresponding AUC–ROC and AUC–PR values for MusiteDeep (**a**,**b**) and SiteTack: (**c**,**d**).

3.3. Analysis of False Positive Predictions

We have additionally analyzed the false PTM positives, or more specifically, the PTM sites that were predicted by all three tools, but were not supported by experimental, high-confidence reference data. Figure 12 shows that SiteTack is the application which tends to produce the highest number of false positives, counting 97 false positives in MYH7, 81 in MYBPC3, 25 in TNNT2, and 23 in TNNI3, while MusiteDeep is the least prone to false positives.

False positives predicted by consensus of all three applications—MusiteDeep, PTMGPT2, and SiteTack—were also analyzed. We have discovered a total of 11 consensus false positives for HTP > 1, of which 7 are phosphorylation sites, 2 acetylation and 2 methylation sites. The phosphorylation sites are: T1513 and S1735 in MYH7, and S18, S297, T307, T705, Y79 in MYBPC3. The acetylation residues are K1451 (MYH7) and K14 (MYBPC3), while the methylation residues are K129 (MYH7) and R44 (MYBPC3). There were no consensus false positives in TNNT2 and TNNI3 proteins. Of these 11 consensus false positives (for HTP > 1), 5 (five) residues are completely absent from the experimentally validated set, regardless of the HTP value. Two consensus false positives are found in MYH7 protein: K1451 (acetylation) and K129 (methylation), and three are found in MYBPC3: T705 (phosphorylation), K14 (acetylation), R44 (methylation), as summarized in Table 5 and visualized in Figure 13.

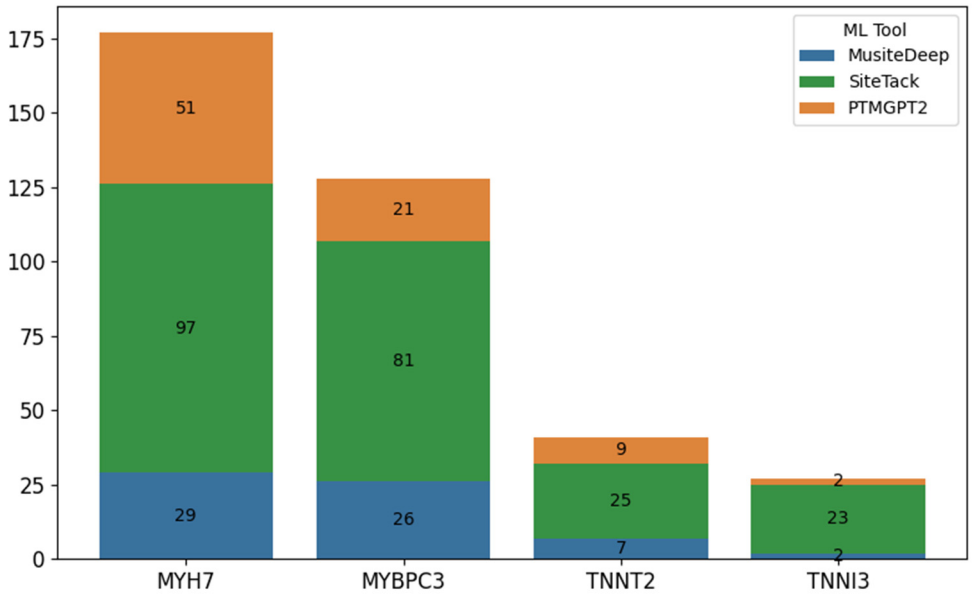


Figure 12. Number of false positives per tool and protein.

Table 5. PTM sites identified in consensus by all three ML tools, which are not experimentally supported at PhosphoSitePlus.

Protein	Phosphorylation	Acetylation	Methylation
MYH7		K1451	K129
MYBPC3	T705	K14	R44

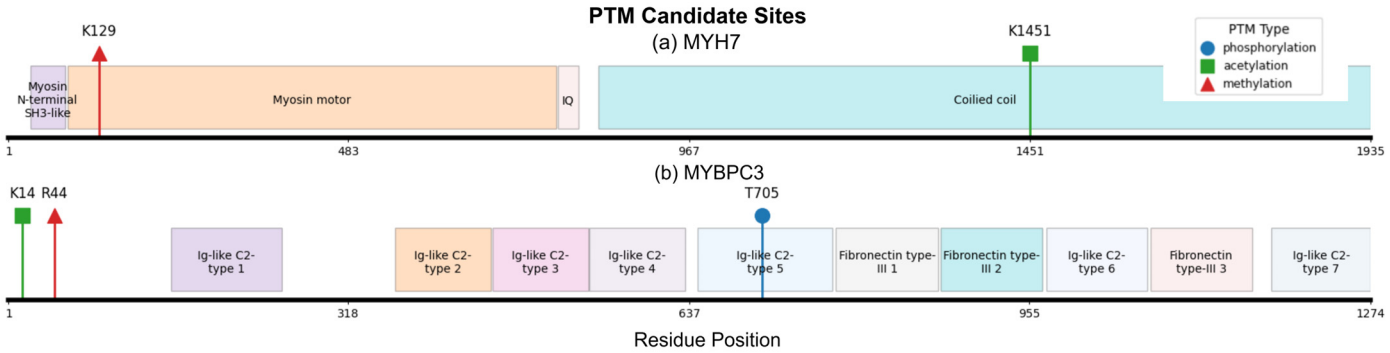


Figure 13. PTM candidate sites identified in consensus for: (**a**) MYH7 protein and (**b**) MYBPC3 protein aligned with protein domain structure.

MYH7 (β-myosin heavy chain) and *MYBPC3* (cardiac myosin-binding protein C) are the two most common sarcomere genes linked to HCM, responsible for ≈50% of familial HCM cases [16]. To investigate the potential functional impact of the five PTM candidates: K129 and K1451 in MYH7 and K14, R44, T705 in MYBPC3 (Table 5), we have analyzed publicly available protein databases, including UniProt, Protein Data Bank (PDB), and AlphaFold Protein Structure Database. MYH17 K1451 is in the coiled-coil tail of the protein (839–1935), involved in thick filament assembly. Acetylation at this particular residue will neutralize lysine charge that can alter thick filament backbone interactions or affect crucial binding mechanisms, such as titin binding or binding to other myosin tails. MYH7 K129 is located in the myosin motor (head) domain (85–778), which binds ATP, interacts with actin, and undergoes large conformational changes during force generation. The domain is also highly conserved. The functional impact of a PTM in this region is expected to be more significant compared to that in the tail. MYBPC3 T705 is placed within the central

Ig/Fn domain, or more precisely, Ig-like C2-type 5 (645–771). Phosphorylation adds a negative charge to the structured domain, which may alter how MYBPC3 interacts with myosin heads or actin, with a possible impact on tension regulation. MYBPC3 K14 is in the N-terminal region, important for binding myosin and modulating actin interaction. Acetylation neutralizes positive charge, which may weaken or shift regulatory binding interactions at the extreme N-terminal, potentially resulting in dysregulation. MYBPC3 R44 is also in the N-terminal region. That could affect protein stability or the local interaction network. All five PTM sites were identified as priority candidates for further experimental validation, given their potential functional roles, which suggest they may significantly impact protein activity and contribute to key biological processes.

3.4. Proposed Experimental Validation

To experimentally validate the predicted post-translational modification (PTM) sites in MYH7 (K1451 acetylation, K129 methylation) and MYBPC3 (T705 phosphorylation, K14 acetylation, R44 methylation), several approaches can be employed. For example, site-directed mutagenesis can generate non-modifiable or PTM-mimetic variants (e.g., MYH7 K1451R/K1451Q or MYBPC3 T705A/T705D), which can be introduced into human-induced pluripotent stem cell-derived cardiomyocytes (iPSC-CMs) to evaluate effects on sarcomere organization, calcium handling, and contractility using functional assays such as engineered heart tissues or traction force microscopy. IPSC-CM models are widely used to reproduce sarcomeric defects and contractile abnormalities observed in hypertrophic cardiomyopathy (HCM) caused by MYH7 or MYBPC3 alterations [27]. Also, targeted proteomics approaches, including parallel reaction monitoring (PRM) or selected reaction monitoring (SRM), can be applied to cardiac tissue samples from HCM patients and controls to detect and quantify peptides containing these modified residues, thereby confirming whether these PTMs occur in vivo. Third, structural and interaction-based studies could be performed by combining AlphaFold structural modeling with biochemical assays, such as co-immunoprecipitation, to determine whether PTM-mimetic mutations alter interactions among MYH7, MYBPC3, and other sarcomeric proteins.

Confirming these PTMs could significantly refine our mechanistic understanding of sarcomere regulation in HCM. For example, phosphorylation of cardiac myosin-binding protein C is known to regulate the availability of myosin heads for actin interaction and influence calcium-dependent force generation, demonstrating how PTMs can directly modulate cross-bridge cycling and contractile dynamics [28]. Similarly, alterations in MYH7 or MYBPC3 interactions within the thick filament can destabilize the super-relaxed state of myosin and increase the number of active cross-bridges, a mechanism thought to contribute to hypercontractility in HCM [29]. Therefore, validating these candidate PTMs could reveal an additional regulatory layer in which dynamic biochemical modifications influence thick filament assembly, cross-bridge availability, and calcium sensitivity, thereby shaping the molecular mechanisms underlying HCM pathophysiology.

4. Discussion

In this study, we have selected three contemporary ML-based tools (MusiteDeep, PTMGPT2, and SiteTack), all accessible as web applications, capable of predicting post-translational modification (PTM) sites across four modification types of interest: phosphorylation, methylation, ubiquitination, and acetylation. We evaluated their performance using four protein sequences associated with HCM: MYH7, MYBPC3, TNNT2, and TNNI3. Although numerous PTM prediction tools have been reported in the literature, most are designed to detect a single modification type—for example, NetPhos [30], DeepNphos [31] for phosphorylation, DeepAcet [32] and DeepUbi [33] for acetylation and ubiquitina-

tion respectively, GPS-MSP for methylation [34]—thereby limiting their applicability for comprehensive multi-PTM analyses. Additionally, there are several tools for predicting residue-targeted PTMs, such as RMTLysPTM [35] for predicting lysine-targeted PTMs: acetylation, crotonylation, methylation, and succinylation, and MUscADEL [36], which is a deep bidirectional LSTM framework designed to predict eight lysine-targeted PTMs in human and mouse proteomes.

Our evaluation was based on predefined thresholds selected prior to testing, ensuring consistent and comparable results. Specifically, it is more than one HTP for PhosphoSitePlus, from where we obtained the experimentally verified PTMs, and a prediction score threshold of 0.6 for the ML tools. The results from MYH7, MYBPC3, TNNT2 and TNNI3, representing a small panel of sarcomeric proteins implicated in HCM, showed that the tools behaved differently across the different PTM types. Phosphorylation was the only PTM type for which all three ML tools could be compared, showing that MusiteDeep has a more conservative prediction, with higher precision but lower recall, while SiteTack favored recall, making it more suitable for scenarios where detecting all PTM sites is more important than avoiding false positives. MusiteDeep could not accurately detect acetylation sites, while PTMGPT2 showed better performance than SiteTack. The latter also showed the ability to identify ubiquitination sites. All three tools failed to identify any true positive methylation sites. However, with a sample size restricted to four proteins, performance metrics exhibit considerable protein-specific variability. Although mean metrics are reported, this variation prevents robust statistical comparison across tools. Consequently, we focus our conclusions on broad, qualitative insights into overall tool behavior.

According to the study by Gutierrez et al., SiteTack + PTMs outperformed MusiteDeep in terms of AUC and Area under the precision–recall curve [23]. Although SiteTack showed better overall performance, MusiteDeep performed better for acetylation and achieved a higher AUC for tyrosine phosphorylation (Y) [23]. In the study where PTMGPT2 was introduced, the tool was compared with MusiteDeep, indicating that it outperformed MusiteDeep in methylation, ubiquitination, acetylation, and phosphorylation(Y), showing better precision, recall, and F1-score [24]. For phosphorylation (S, T) PTMPGPT2 showed better precision and F1-score, but MusiteDeep had better MCC and recall [24].

The direct primary use of MusiteDeep, SiteTack, and PTMGPT as central PTM analysis tools remains underrepresented in the literature, resulting in a lack of comparable results across additional studies. MusiteDeep, on the other hand, is widely referenced in later papers as a benchmark model in PTM prediction research, especially in studies proposing new deep learning methods [37].

We additionally explored false positives consistently predicted by the three tools, identifying five residues not experimentally supported by PhosphoSitePlus: K1451 (acetylation) and K129 (methylation) for MYH7, T705 (phosphorylation) for MYBPC3, and K14 (acetylation) and R44 (methylation) for MYBPC3. Given that PTMs play essential roles in regulating sarcomere contractility, they can significantly influence disease severity, progression, and phenotypic manifestation. Therefore, the identified high-priority candidates present potential targets for future biological exploration.

Finally, it is important to explicitly acknowledge that the absence of a PTM annotation in PhosphoSite (used as a baseline) does not imply true absence in vivo, particularly for less-studied PTM types or poorly characterized proteins. Consequently, the reported performance metrics and false positives are conditional on this specific reference dataset and may underestimate the performance of tools that correctly predict biologically real but not yet annotated modification sites.

5. Conclusions

In this study, we have analyzed the ability of three machine learning tools (MusiteDeep, PTMGPT2, and SiteTack) to predict phosphorylation, acetylation, methylation, and ubiquitination sites on a very specific, small panel of sarcomeric proteins. The proteins were selected for their biological relevance to HCM, enabling us to evaluate tool performance in a focused, disease-relevant context. However, because our study is focused on a restricted protein set, the analysis may not reflect the full predictive capabilities of the tools. We have analyzed the number of detected PTMs under different predicting score thresholds and evaluated their performances, using precision, recall and F1-score. The results showed that MusiteDeep can be used when a precise detection of phosphorylation is needed for the major contributing HCM-associated proteins. For the same target proteins, PTMGPT2 showed best performances for acetylation. On the other hand, SiteTack can be suitable for phosphorylation site screening and for exploring other PTM types in HCM-related proteins, but it tends to produce the highest number of false positives. Finally, we report on five unknown PTM sites in the MYH7 and MYBPC3 proteins, suggesting them as high-priority candidates for experimental validation. Although these modifications have not yet been confirmed experimentally, their predicted functional impact makes them promising targets for future studies aimed at understanding sarcomeric regulation in HCM. Future functional assays will be critical to determine their ultimate PTM status.

Author Contributions: Conceptualization, D.S.; methodology, D.S. and N.T.; software, N.T.; validation, D.S.; formal analysis, N.T.; investigation, N.T. and L.J.; resources, N.T.; data curation, N.T.; writing—original draft preparation, N.T., L.J. and D.S.; writing—review and editing, N.T., L.J. and D.S.; visualization, N.T. and L.J.; supervision, D.S.; project administration, D.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available from the author, [N.T.], upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Mann, M.; Jensen, O.N. Proteomic analysis of post-translational modifications. *Nat. Biotechnol.* **2003**, *21*, 255–261. [CrossRef]
2. Ramazi, S.; Zahiri, J. Posttranslational modifications in proteins: Resources, tools and prediction methods. *Database J. Biol. Databases Curation* **2021**, *2021*, baab012. [CrossRef]
3. Greenblatt, S.M.; Frye, R.J.M. Protein arginine methylation: From enigmatic functions to therapeutic targeting. *Nat. Rev. Drug Discov.* **2021**, *20*, 509–530. [CrossRef]
4. Bradley, D. The evolution of post-translational modifications. *Curr. Opin. Genet. Dev.* **2022**, *76*, 101956. [CrossRef] [PubMed]
5. Kokot, T.; Köhn, M. Emerging insights into serine/threonine-specific phosphoprotein phosphatase function and selectivity. *J. Cell Sci.* **2022**, *135*, jcs259618. [CrossRef]
6. Zhong, Q.; Xiao, X.; Qiu, Y.; Xu, Z.; Chen, C.; Chong, B.; Zhao, X.; Hai, S.; Li, S.; An, Z.; et al. Protein posttranslational modifications in health and diseases: Functions, regulatory mechanisms, and therapeutic implications. *MedComm* **2023**, *4*, e261. [CrossRef]
7. Kumar, M.; Haghighi, K.; Kranias, E.G.; Sadayappan, S. Phosphorylation of cardiac myosin-binding protein-C contributes to calcium homeostasis. *J. Biol. Chem.* **2020**, *295*, 11275–11291. [CrossRef] [PubMed]
8. Zhang, H.; Guo, H.; Han, F.; Zheng, Y. Regulatory mechanisms of m6A methylation in dilated cardiomyopathy. *Am. J. Transl. Res.* **2025**, *17*, 47–59. [CrossRef]
9. Menzies, K.J.; Zhang, H.; Katsyuba, E.; Auwerx, J. Protein acetylation in metabolism—Metabolites and cofactors. *Nat. Rev. Endocrinol.* **2016**, *12*, 43–60. [CrossRef]

10. Li, Z.; Chen, J.; Huang, H.; Zhan, Q.; Wang, F.; Chen, Z.; Lu, X.; Sun, G. Post-translational modifications in diabetic cardiomyopathy. *J. Cell. Mol. Med.* **2024**, *28*, e18158. [CrossRef]
11. Neutzner, M.; Neutzner, A. Enzymes of ubiquitination and deubiquitination. *Essays Biochem.* **2012**, *52*, 37–50. [CrossRef]
12. Guo, H.J.; Rahimi, N.; Tadi, P. Biochemistry, ubiquitination. In *StatPearls [Internet]*; StatPearls Publishing: Treasure Island, FL, USA, 2023. Available online: https://www.ncbi.nlm.nih.gov/books/NBK556052/ (accessed on 5 September 2025).
13. Day, S.M. The ubiquitin proteasome system in human cardiomyopathies and heart failure. *Am. J. Physiol. Heart Circ. Physiol.* **2013**, *304*, H1283–H1293. [CrossRef]
14. Maron, B.J.; Maron, M.S. Hypertrophic cardiomyopathy. *Lancet* **2013**, *381*, 242–255. [CrossRef] [PubMed]
15. De Gaspari, M.; Basso, C.; Perazzolo Marra, M.; Elia, S.; Bueno Marinas, M.; Angelini, A.; Thiene, G.; Rizzo, S. Small Vessel Disease: Another Component of the Hypertrophic Cardiomyopathy Phenotype Not Necessarily Associated with Fibrosis. *J. Clin. Med.* **2021**, *10*, 575. [CrossRef] [PubMed]
16. Marian, A.J.; Braunwald, E. Hypertrophic cardiomyopathy: Genetics, pathogenesis, clinical manifestations, diagnosis, and therapy. *Circ. Res.* **2017**, *121*, 749–770. [CrossRef] [PubMed]
17. Teekakirikul, P.; Zhu, W.; Huang, H.C.; Fung, E. Hypertrophic Cardiomyopathy: An Overview of Genetics and Management. *Biomolecules* **2019**, *9*, 878. [CrossRef]
18. Tucholski, T.; Cai, W.; Gregorich, Z.R.; Bayne, E.F.; Mitchell, S.D.; McIlwain, S.J.; de Lange, W.J.; Wrobbel, M.; Karp, H.; Hite, Z.; et al. Distinct hypertrophic cardiomyopathy genotypes result in convergent sarcomeric proteoform profiles revealed by top-down proteomics. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 24691–24700. [CrossRef]
19. Lopes, L.R.; Ho, C.Y.; Elliott, P.M. Genetics of hypertrophic cardiomyopathy: Established and emerging implications for clinical practice. *Eur. Heart J.* **2024**, *45*, 2727–2734. [CrossRef]
20. Huang, X.; Feng, Z.; Liu, D.; Gou, Y.; Chen, M.; Tang, D.; Han, C.; Peng, J.; Peng, D.; Xue, Y. PTMD 2.0: An updated database of disease-associated post-translational modifications. *Nucleic Acids Res.* **2025**, *53*, D554–D563. [CrossRef]
21. Crocini, C.; Gotthardt, M. Cardiac sarcomere mechanics in health and disease. *Biophys. Rev.* **2021**, *13*, 637–652. [CrossRef]
22. Wang, D.; Liu, D.; Yuchi, J.; He, F.; Jiang, Y.; Cai, S.; Li, J.; Xu, D. MusiteDeep: A deep-learning based webserver for protein post-translational modification site prediction and visualization. *Nucleic Acids Res.* **2020**, *48*, W140–W146. [CrossRef]
23. Gutierrez, C.S.; Kassim, A.A.; Gutierrez, B.D.; Raines, R.T. Sitetack: A deep learning model that improves PTM prediction by using known PTMs. *Bioinformatics* **2024**, *40*, btae602. [CrossRef] [PubMed]
24. Shrestha, P.; Kandel, J.; Tayara, H.; Chong, K.T. Post-translational modification prediction via prompt-based fine-tuning of a GPT-2 model. *Nat. Commun.* **2024**, *15*, 6699. [CrossRef]
25. Hornbeck, P.V.; Zhang, B.; Murray, B.; Kornhauser, J.M.; Latham, V.; Skrzypek, E. PhosphoSitePlus, 2014: Mutations, PTMs and recalibrations. *Nucleic Acids Res.* **2015**, *43*, D512–D520. [CrossRef]
26. Hornbeck, P.V.; Kornhauser, J.M.; Latham, V.; Murray, B.; Nandhikonda, V.; Nord, A.; Skrzypek, E.; Wheeler, T.; Zhang, B.; Gnad, F. 15 years of PhosphoSitePlus®: Integrating post-translationally modified sites, disease variants and isoforms. *Nucleic Acids Res.* **2019**, *47*, D433–D441. [CrossRef]
27. Birket, M.J.; Ribeiro, M.C.; Kosmidis, G.; Ward, D.; Leitoguinho, A.R.; van de Pol, V.; Dambrot, C.; Devalla, H.D.; Davis, R.P.; Mastroberardino, P.G.; et al. Contractile Defect Caused by Mutation in MYBPC3 Revealed under Conditions Optimized for Human PSC-Cardiomyocyte Function. *Cell Rep.* **2015**, *13*, 733–745. [CrossRef]
28. Dutsch, A.; Wijnker, P.J.M.; Schlossarek, S.; Friedrich, F.W.; Krämer, E.; Braren, I.; Hirt, M.N.; Brenière-Letuffe, D.; Rhoden, A.; Mannhardt, I.; et al. Phosphomimetic cardiac myosin-binding protein C partially rescues a cardiomyopathy phenotype in murine engineered heart tissue. *Sci. Rep.* **2019**, *9*, 18152. [CrossRef]
29. Spudich, J.A.; Nandwani, N.; Robert-Paganin, J.; Houdusse, A.; Ruppel, K.M. Reassessing the unifying hypothesis for hypercontractility caused by myosin mutations in hypertrophic cardiomyopathy. *EMBO J.* **2024**, *43*, 4139–4155. [CrossRef] [PubMed]
30. Blom, N.; Gammeltoft, S.; Brunak, S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J. Mol. Biol.* **1999**, *294*, 1351–1362. [CrossRef]
31. Chang, X.; Zhu, Y.; Chen, Y.; Li, L. DeepNphos: A deep-learning architecture for prediction of N-phosphorylation sites. *Comput. Biol. Med.* **2024**, *170*, 108079. [CrossRef] [PubMed]
32. Wu, M.; Yang, Y.; Wang, H.; Xu, Y. A deep learning method to more accurately recall known lysine acetylation sites. *BMC Bioinform.* **2019**, *20*, 49. [CrossRef] [PubMed]
33. Fu, H.; Yang, Y.; Wang, X.; Wang, H.; Xu, Y. DeepUbi: A deep learning framework for prediction of ubiquitination sites in proteins. *BMC Bioinform.* **2019**, *20*, 86. [CrossRef] [PubMed]
34. Deng, W.; Wang, Y.; Ma, L.; Zhang, Y.; Ullah, S.; Xue, Y. Computational prediction of methylation types of covalently modified lysine and arginine residues in proteins. *Brief. Bioinform.* **2017**, *18*, 647–658. [CrossRef]
35. Chen, L.; Chen, Y. RMTLysPTM: Recognizing multiple types of lysine PTM sites by deep analysis on sequences. *Brief. Bioinform.* **2023**, *25*, bbad450. [CrossRef]

36. Chen, Z.; Liu, X.; Li, F.; Li, C.; Marquez-Lago, T.; Leier, A.; Akutsu, T.; Webb, G.I.; Xu, D.; Smith, A.I.; et al. Large-scale comparative assessment of computational predictors for lysine post-translational modification sites. *Brief. Bioinform.* **2019**, *20*, 2267–2290. [CrossRef]
37. Han, Y.; He, F.; Shao, Q.; Wang, D.; Xu, D. MTPrompt-PTM: A Multi-Task Method for Post-Translational Modification Prediction Using Prompt Tuning on a Structure-Aware Protein Language Model. *Biomolecules* **2025**, *15*, 843. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.