PNAS RESEARCH ARTICLE | BIOCHEMISTRY CHEMISTRY

High sensitivity top–down proteomics captures single muscle cell heterogeneity in large proteoforms

Jake A. Melby[a], Kyle A. Brown[a], Zachery R. Gregorich[b], David S. Roberts[a], Emily A. Chapman[a], Lauren E. Ehlers[a], Zhan Gao[c], Eli J. Larson[a], Yutong Jin[a], Justin R. Lopez[d], Jared Hartung[d], Yanlong Zhu[c,e], Sean J. McIlwain[f], Daojing Wang[g], Wei Guo[b], Gary M. Diffee[d], and Ying Ge[a,c,e,1]

Edited by Catherine Murphy, University of Illinois at Urbana-Champaign, Urbana, IL; received December 31, 2022; accepted April 5, 2023

Single-cell proteomics has emerged as a powerful method to characterize cellular phenotypic heterogeneity and the cell-specific functional networks underlying biological processes. However, significant challenges remain in single-cell proteomics for the analysis of proteoforms arising from genetic mutations, alternative splicing, and post-translational modifications. Herein, we have developed a highly sensitive functionally integrated top–down proteomics method for the comprehensive analysis of proteoforms from single cells. We applied this method to single muscle fibers (SMFs) to resolve their heterogeneous functional and proteomic properties at the single-cell level. Notably, we have detected single-cell heterogeneity in large proteoforms (>200 kDa) from the SMFs. Using SMFs obtained from three functionally distinct muscles, we found fiber-to-fiber heterogeneity among the sarcomeric proteoforms which can be related to the functional heterogeneity. Importantly, we detected multiple isoforms of myosin heavy chain (~223 kDa), a motor protein that drives muscle contraction, with high reproducibility to enable the classification of individual fiber types. This study reveals single muscle cell heterogeneity in large proteoforms and establishes a direct relationship between sarcomeric proteoforms and muscle fiber types, highlighting the potential of top–down proteomics for uncovering the molecular underpinnings of cell-to-cell variation in complex systems.

proteomics | mass spectrometry | single cell | proteoform | single muscle fiber

Significance

Single-cell technologies are revolutionizing biology and molecular medicine by allowing direct investigation of the biological variability among individual cells. Top–down proteomics is uniquely capable of dissecting biological heterogeneity at the intact protein level. Herein, we develop a highly sensitive single-cell top–down proteomics method to reveal diverse molecular variations in large proteins (>200 kDa) among individual single muscle cells. Our results both reveal and characterize the differences in protein post-translational modifications and isoform expression possible between individual muscle cells. We further integrate functional properties with proteomics and reproducibly measure myosin isoforms for individual muscle fiber type classification. Our study highlights the potential of top–down proteomics for understanding how single-cell protein heterogeneity contributes to cellular functions.

Single-cell analysis has revealed that even morphologically and genetically identical cells can differ dramatically in their functional properties, as well as how they respond to stressors and extracellular ques (1–4). Mass spectrometry (MS)-based methods for single-cell proteomic analysis hold great promise for unraveling the molecular underpinnings of cellular heterogeneity and thus are highly desired (4–12) but have lagged behind the development of single-cell genomics and transcriptomics (3, 13, 14). The major challenges for MS-based single-cell proteomics are the limited protein content of most mammalian cell types, the broad dynamic range of the proteome, and the high complexity in proteoforms arising from sequence variations and posttranslational modifications (PTMs) (4, 6, 12, 15).

The two principal MS-based proteomics techniques are the peptide-centric "bottom–up" and protein-centric "top–down" approaches (16–18). Since peptides are more easily separated, ionized, and fragmented than intact proteins, bottom–up proteomics has served as the workhorse for MS-based proteomics (18–20). Nearly all MS-based single-cell proteomics studies to date employ bottom–up proteomics (8–11, 21, 22). Nevertheless, this approach suffers from intrinsic limitations, including the protein inference problem and loss of information and connectivity in mapping sequence variations and PTMs (23, 24). Top–down proteomics circumvents these limitations by analyzing intact proteins directly allowing for the identification and analysis of proteoforms—a term encompassing the myriad protein products arising from a single gene as a result of sequence variations, alternative splicing, and PTMs (24–27). Although inherently less sensitive than current bottom–up approaches due to the exponential decay in the signal-to-noise (S/N) ratio with increasing molecular weight (28), top–down proteomics is ideally suited for dissecting cellular heterogeneity at the level of proteoforms (17, 24–27, 29–31).

In this study, we developed a high-sensitivity top–down proteomics method to examine proteoform heterogeneity at the single-cell level. We applied this method to single muscle fiber (SMFs), multinucleated single muscle cells with heterogeneous structural and functional properties (32–35), to establish a direct relationship between proteoforms and muscle fiber types (Fig. 1 *A* and *B*). Notably, we have detected large proteoforms (>200 kDa) and visualized the single-cell heterogeneity in the SMFs. Using SMFs obtained from three functionally distinct muscle types, we found fiber-to-fiber heterogeneity in a large number of proteoforms in the sarcomere. Sarcomere is the basic contractile units in

Competing interest statement: A co-author, Dr. Daojing Wang, is an employee of Newomics, owns the stocks of Newomics, and is the founder and CEO of Newomics.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: ying.ge@wisc.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2222081120/-/DCSupplemental.

Published May 1, 2023.

muscle, which consists of thin and thick filaments flanked serially by dense protein structures known as Z-disks (36–39). We have simultaneously characterized PTMs together with isoforms (from different genes) from SMFs. Importantly, we detected multiple isoforms of myosin heavy chain (MyHC, 223 kDa) (34, 35, 40–43), a motor protein of thick filament playing a critical role in muscle contraction, from the SMFs with high reproducibility enabling the classification of fiber type at the single-cell resolution. As illustrated, top–down proteomics of SMFs provides single-cell resolution and reveals biological phenomena that are masked by bulk proteomics analysis.

Results

A High Sensitivity Top–Down Proteomics Platform. Our top–down proteomics method is designed to minimize adsorptive protein losses by performing the protein extraction and follow-up sample processing in one pot using MS-compatible solvents. Subsequently, the extracted proteins are separated using low-flow capillary liquid chromatography (LC) coupled to a microflow multiemitter nanoelectrospray (MnESI) source to increase ionization efficiency followed by tandem MS (MS/MS) for intact protein analysis (Fig. 1).

Inspired by previous bottom–up single-cell proteomic studies (9, 10, 44–46), we sought to establish a one-pot sample preparation method to minimize protein losses (Fig. 1*C* and *SI Appendix*, Fig. S1). First, we optimized the lysis buffers to effectively extract proteins using a minimal amount of tissue samples (1 mg) (*SI Appendix*, *Supplemental Note 1* and Fig. S2). Addition of hexafluoroisopropanol (HFIP) proved to be the most effective at extracting proteins from a small amount of tissue sample while maintaining MS compatibility (*SI Appendix*, Fig. S2). Subsequently, by evaluation of different lysis techniques, we found that freeze–thaw lysis clearly performed better than sonication for protein extraction from SMFs (*SI Appendix*, *Supplemental Note 2* and Fig. S3). Next, we assessed various concentrations of HFIP in the extraction solution and chose 25% HFIP which yielded the greatest protein extraction and maintained MS compatibility (*SI Appendix*, *Supplemental Note 3* and Fig. S4). To determine whether freeze–thaw lysis in HFIP affected protein PTMs (47), we compared this method to our established protein extraction protocol (48, 49) (*SI Appendix*, *Supplemental Note 4*). Our results have shown no discernible changes in PTMs or proteolysis resulting from freeze–thaw lysis (*SI Appendix*, Fig. S5). Thus, we have chosen a solution of 25% HFIP and a freeze–thaw cycle for cell lysis and protein extraction in this one-pot sample preparation method (Fig. 1*C*).

We next developed a highly sensitive top–down LC-MS/MS method by minimizing sample dilution during reversed-phased liquid chromatography (RPLC) and maximizing ionization efficiency. We chose a capillary RPLC column (Thermo MAbPac) with a narrow column inner diameter (150 μm) to minimize sample dilution during protein separation and provide robust separation afforded by μL flow rates (2 μL/min). The capillary column was coupled to a Newomics micro-flow-nanospray electrospray ionization (MnESI) source, which splits μL flow rates into 8 nanoelectrospray ionization (nanoESI) emitters substantially increasing ionization efficiency (50, 51) (*SI Appendix*, Fig. S6). The MnESI source allows for high sensitivity, throughput, and quantitative accuracy for intact protein characterization (52, 53) that resulted in a 0.5-ng detection limit of a standard protein using our method (*SI Appendix*, Fig. S6). When performing a targeted MS/MS experiment of a 10-ng injection of carbonic anhydrase (*SI Appendix*,

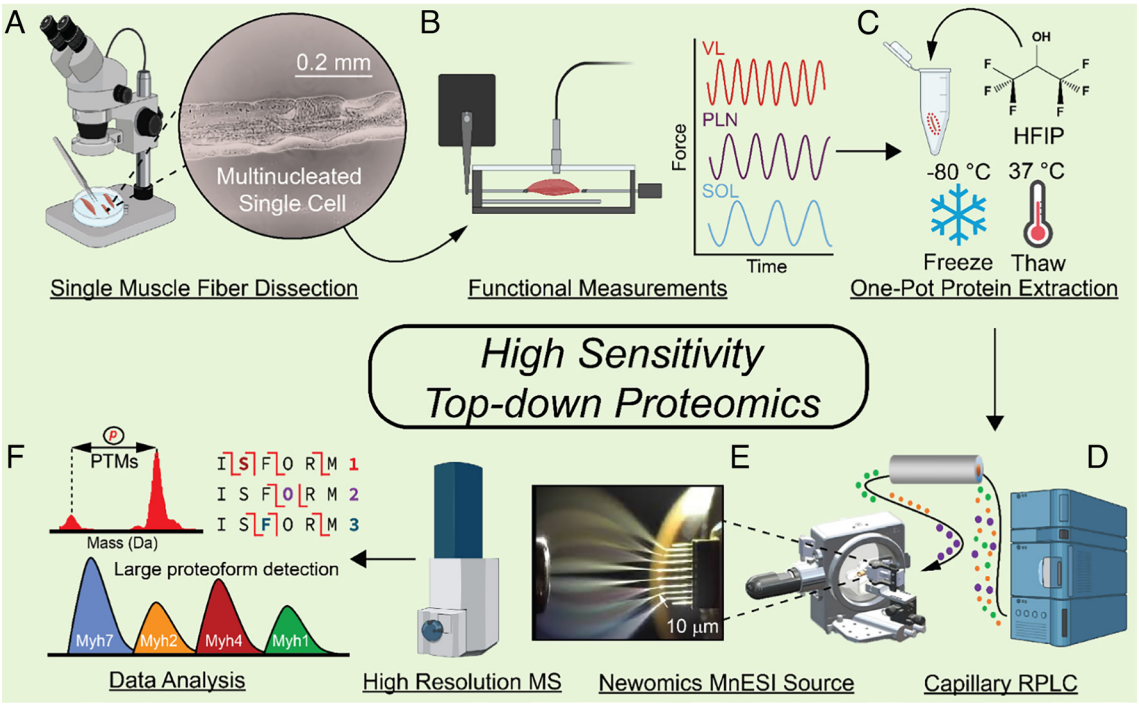

Fig. 1. High-sensitivity top–down proteomics of single muscle fibers (SMFs), multinucleated single muscle cells. (*A*) SMFs were mechanically dissociated from their constituent muscles in relaxation buffer using fine-point tweezers and a microscope to visualize the fibers. (*B*) SMFs from each muscle were placed in a force transducer to measure their shortening velocity (n = 10 fibers per muscle). (*C*) A separate group of SMFs were placed directly into individual low protein binding microcentrifuge tubes for top–down proteomics measurements (n = 6 fibers per muscle). Proteins were extracted from SMFs using hexafluoro-2-propanol (HFIP) and a freeze–thaw lysis. (*D*) Proteins from SMF extracts were loaded onto a capillary reversed phased liquid chromatography (RPLC) column and separated based on their hydrophobicity. (*E*) Eluting proteins from SMFs were ionized with a Newomics microflow nanospray electrospray ion (MnESI) source and analyzed using a Bruker maXis II mass spectrometer. (*F*) Analysis of the resultant top–down proteomics data included posttranslational modification (PTM) measurements, isoform characterization using tandem mass spectrometry (MS/MS), as well as detection of large proteoforms.

Supplemental Note 5), we found that there was a two-fold gain in MS1 S/N and almost twice as many fragment ions detected by using the Newomics MnESI source compared to a conventional ESI source (*SI Appendix*, Fig. S7). Proteins elute earlier in the MnESI source due to smaller dead volume in the source (52, 53), which limits the chromatographic peak dispersion that results in sharper chromatographic peaks. We combined this top–down proteomics method with one-pot sample preparation to achieve minimal sample losses and highly sensitive capillary LC-MS/MS for proteoforms (from the same gene) and isoforms (from different genes of the same protein family) characterization (Fig. 1 *E*–*G*).

Integration of Functional Properties and Top–down Proteomic Analysis of SMFs. Next, we applied our top–down proteomics approach to SMFs (multinucleated single muscle cells) because skeletal muscle is remarkably heterogeneous, containing a mixture of SMFs, blood vessels, nerves, and connective tissue (32–35, 39, 42) (Fig. 2 *A* and *B*); thus, methods to study contractile proteoforms at the single fiber level are highly desirable. SMFs are broadly classified as either fast-twitch or slow-twitch based on their contractile and metabolic properties (35). We chose the vastus lateralis (VL), plantaris (PLN), and soleus (SOL) muscles because VL and SOL contain predominantly fast- and slow-twitch fibers, respectively, whereas PLN contains a mixture of fast- and slow-twitch fibers (35, 54) (Figs. 2*A*).

To determine the unique functional properties amongst fiber types from three muscle types (VL, PLN, SOL), SMFs from each muscle group (n = 10) were dissected in a relaxation buffer that contained protease and phosphatase inhibitors, as well as antioxidants, under a confocal microscope (Fig. 1*B* and *SI Appendix*).

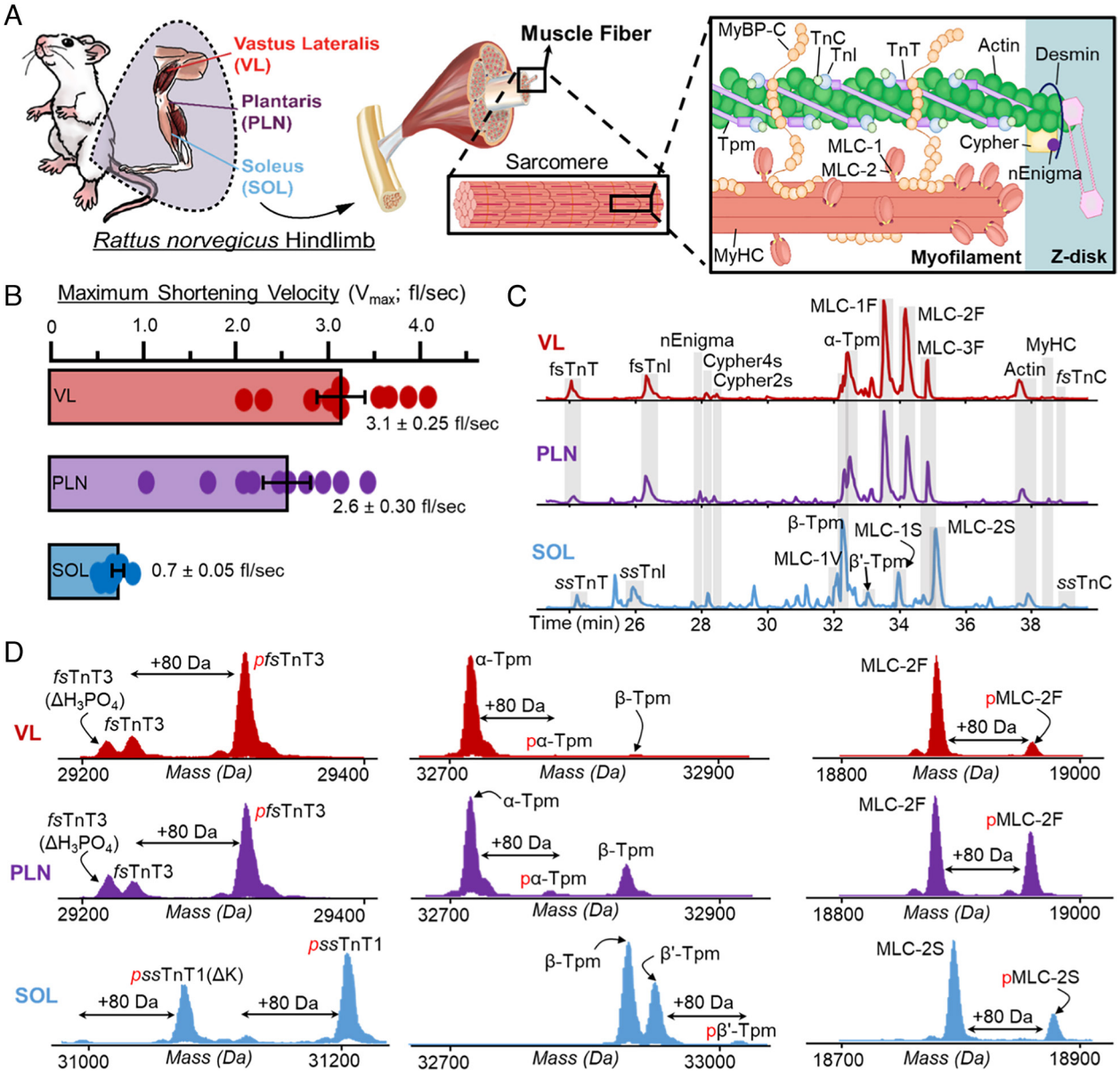

Fig. 2. SMFs obtained from skeletal muscles have unique contractile properties and sarcomere proteoform landscapes. (*A*) Schematic representation of skeletal muscle structure, which are made up of SMFs. Within SMFs are the sarcomeres, which contain the proteins necessary for contraction and relaxation including thin filament proteins (TnT, TnI, TnC, Tpm, and Actin), thick filament proteins (MLC-1, MLC-2, MyHC), and Z-disk proteins (e.g., Cypher2s, Cypher 4s, and nEnigma). (*B*) Measurement of the maximum shortening velocity (V_{max}; fiber lengths (fl)/s) of SMFs isolated from the VL, PLN, and SOL muscles (n = 10 fibers per muscle). (*C*) Separation and detection of major sarcomeric proteoforms detected in SMFs from VL, PLN, and SOL muscles via LC-MS/MS; base peak chromatograms (BPCs) shown (n = 6 fibers per muscle). (*D*) Representative deconvoluted mass spectra displaying proteoforms from VL (fast-twitch), PLN (mixed fast- and slow-twitch), and SOL (slow-twitch) single fibers. Monophosphorylation is indicated by red "p", "ΔH_3PO_4" indicates a loss of phosphate from pfsTnT3, and "-K" indicates loss of a lysine residue from ssTnT1 or pssTnT1.

The average maximum shortening velocities were measured for SMFs obtained from VL [3.23 ± 0.25 fiber lengths/sec (fl/s)], PLN (2.56 ± 0.30 fl/s), and SOL (0.71 ± 0.05 fl/s) muscles (Figs. 1*C* and 2*B* and *SI Appendix*). These values are in agreement with those previously reported in the literature for fast- and slow-twitch muscle fibers isolated from rat muscle (55). There was a larger amount of heterogeneity in the shortening velocity measurements of SMFs from the fast-twitch VL and PLN muscles, which have three distinct fast muscle fiber types (types IIa, IIb, and IIx), compared to the slow-twitch SOL muscle, which only has one slow muscle fiber type (Fig. 2*B*).

Next, we employed top–down proteomics to decipher the proteoform heterogeneities that correlate with unique functional properties in SMFs from VL, PLN, and SOL (Figs. 1 *D*–*F* and 2*C*). The basic contractile apparatus of SMFs is the sarcomere, which consists of thin and thick filaments flanked serially by dense protein structures known as Z-disks (36–38, 56) (Fig. 2*A*). Single muscle cell top–down proteomic analysis of SMFs was highly reproducible across technical replicates (*SI Appendix*, Fig. S8), allowing for the relative quantification of sarcomeric proteoforms between SMFs obtained from the VL, PLN, and SOL muscles (*SI Appendix*, Figs. S9–S11). Several key myofilament (including thin and thick filaments) proteoforms and isoforms were detected in SMFs from the VL, PLN, and SOL muscles (Fig. 2*C* and *SI Appendix*, Table S1). As expected, fast skeletal myofilament protein isoforms, including fast skeletal troponin complex (fsTnT, fsTnI, and fsTnC), alpha tropomyosin (α-Tpm), and fast skeletal myosin light chains (MLC-1F, MLC-2F, and MLC-3F) predominated in fibers from the VL and PLN muscles (Fig. 2 *C* and *D* and *SI Appendix*, Figs. S12–S18). In contrast, slow skeletal troponin complex (ssTnT, ssTnI, and ssTnC), beta-tropomyosin (β-Tpm), and slow skeletal myosin light chains (MLC-1S, MLC-1V, and MLC-2S) were detected in SMFs from SOL muscles (Fig. 2 *C* and *D* and *SI Appendix*, Figs. S12–S19). Considerably, the proteomics results of the isoform distributions can be related to the unique functional properties of the SMFs from VL, PLN, and SOL muscles. Other key myofilament proteins found across all samples included alpha skeletal actin (α-sActin) and MyHC isoforms (*SI Appendix*, Figs. S20 and S21). Additionally, the Z-disk proteins nEnigma (38), Cypher2s, and Cypher4s were found across all fiber samples with varying relative expression (*SI Appendix*, Figs. S22–S24). These important sarcomere proteins were all identified based on their intact protein mass within a 10 ppm mass error, and identifications were further confirmed based on MS/MS data, consistent with previous studies on skeletal muscle top–down proteomics (37, 56, 57). Notably, we were able to measure proteoforms >30 kDa such as TnT isoforms, Cypher isoforms, α-sActin (42 kDa), and MyHC isoforms (223 kDa) (*SI Appendix*, Figs. S12 and S20–S23). These examples illustrated the capability of our top–down proteomics method in detecting intact sarcomeric proteoforms at the level of single muscle cells.

Measurement of Intact MyHC Isoforms for Fiber-Type Classification. Diversity of MyHC isoforms, molecular motor proteins that generate energy for force and contraction, influences the structural and functional properties of SMFs (34, 35, 40–42). MyHC isoforms are approximately 223 kDa in size and share over 80% sequence homology, making them difficult to detect and quantify using bottom–up proteomics, but each isoform corresponds to unique fiber contractile properties (34, 35, 40–42). In rat skeletal muscle, there are four major MyHC isoforms encoded by the genes *Myh1*, *Myh2*, *Myh4*, and *Myh7* (54). MyHC7 is the slow-twitch isoform that is the predominant isoform in SOL tissue (type I), MyHC4 is a fast-twitch isoform that is the predominant isoform found in VL fibers (type IIb), and PLN tissue mainly contains a heterogeneous mixture of the three fast-twitch MyHC isoforms (MyHC2—type IIa, MyHC4—type IIb, and MyHC1—type IIx) (35, 54).

We have observed the characteristic charge state envelope of a large proteoform, as evident by the myriad high charge state ions, at the later phase of 1D RPLC gradient (38 min), unexpectedly (Fig. 3*A* and *SI Appendix*, Fig. S20*A*). The deconvoluted mass spectra revealed that there were four unique masses in total from the SMFs obtained from VL, PLN, and SOL, presumably corresponding to the four major MyHC isoforms found in rat (Fig. 3*B* and *SI Appendix*, Fig. S20*B*). A single MyHC isoform with a mass of 222,679 ± 1 Da was detected in SMFs from SOL, which we putatively assigned as MyHC7 (encoded by *Myh7*) due to the prevalence of type I fibers in this muscle (Fig. 3*C*). Similarly, a single MyHC isoform with a mass of 222,867 ± 2 Da was detected in SMFs from VL, which likely corresponds to MyHC4 (encoded by *Myh4*) based on the high abundance of type IIb fibers in this muscle type (Fig. 3*D*).

MyHC isoform expression was particularly heterogeneous in SMFs obtained from PLN muscle. Three of the SMFs obtained from PLN exhibited dual fiber characteristics, expressing both MyHC4 (type IIb) and a second isoform with a mass of 222,735 ± 1 Da that we putatively assigned MyHC2 (encoded by *Myh2*; type IIa) (Fig. 3*E*). The other three SMFs obtained from PLN expressed MyHC isoforms with a mass of 223,167 ± 1 Da, which we putatively assigned as MyHC1 (type IIx) (Fig. 3*E*). Remarkably, the intact mass analysis of MyHC isoforms was highly consistent across the fiber samples with SDs of 1 to 2 Da, facilitating the classification of SMFs based on MyHC isoform expression. The greater variability in maximum shortening velocities in fibers from PLN muscle reflects the fact that multiple MyHC isoforms are expressed in PLN fibers versus only predominantly one MyHC isoform in SOL and VL fibers (Figs. 2*B* and 3 *D* and *E*) (54, 55). The detection of intact MyHC isoforms at the single muscle cell level was made possible by our high-sensitivity method.

Top–Down MS Analysis of Isoforms Together with PTMs in SMF. Both isoforms and PTMs are known to serve as key regulators for muscle contraction and relaxation (17, 57, 58). We have previously shown that top–down proteomics presents unique advantages in characterizing PTMs together with isoforms encoded by different genes from a multigene family, which often exhibit high sequence homology (37, 38, 59). Here, top–down MS data provide a "bird's eye" view of the sarcomeric proteoforms (arising from PTMs and sequence variations from a single gene) together with isoforms (resulting from different genes in a protein family) (Fig. 2*D*). Many key sarcomere proteins present a diversity of isoforms in three different muscle types (VL, PLN, and SOL) together with PTMs such as N-terminal di-methylation and acetylation, as well as phosphorylation and methylation (Fig. 2*D* and *SI Appendix*, Figs. S12–S24).

Perhaps one of the best examples of using top–down proteomics to analyze PTMs together with isoforms is illustrated by the vast number of fsTnT, the Tpm-binding subunit of the troponin complex, detected in VL and PLN SMFs (Fig. 4). The mammalian *Tnnt3* gene, which encodes fsTnT3, contains 19 exons and can theoretically generate up to 256 variants due to the alternative splicing (60). Additionally, many fsTnT isoforms share high sequence homology and are found in low abundance making their detection by bottom–up proteomics challenging. Lastly, many fsTnT isoforms are known to be posttranslationally modified, such as N-terminal acetylation and phosphorylation (37, 38, 49, 56, 57), adding to the complexity of fsTnT analysis. SMFs from

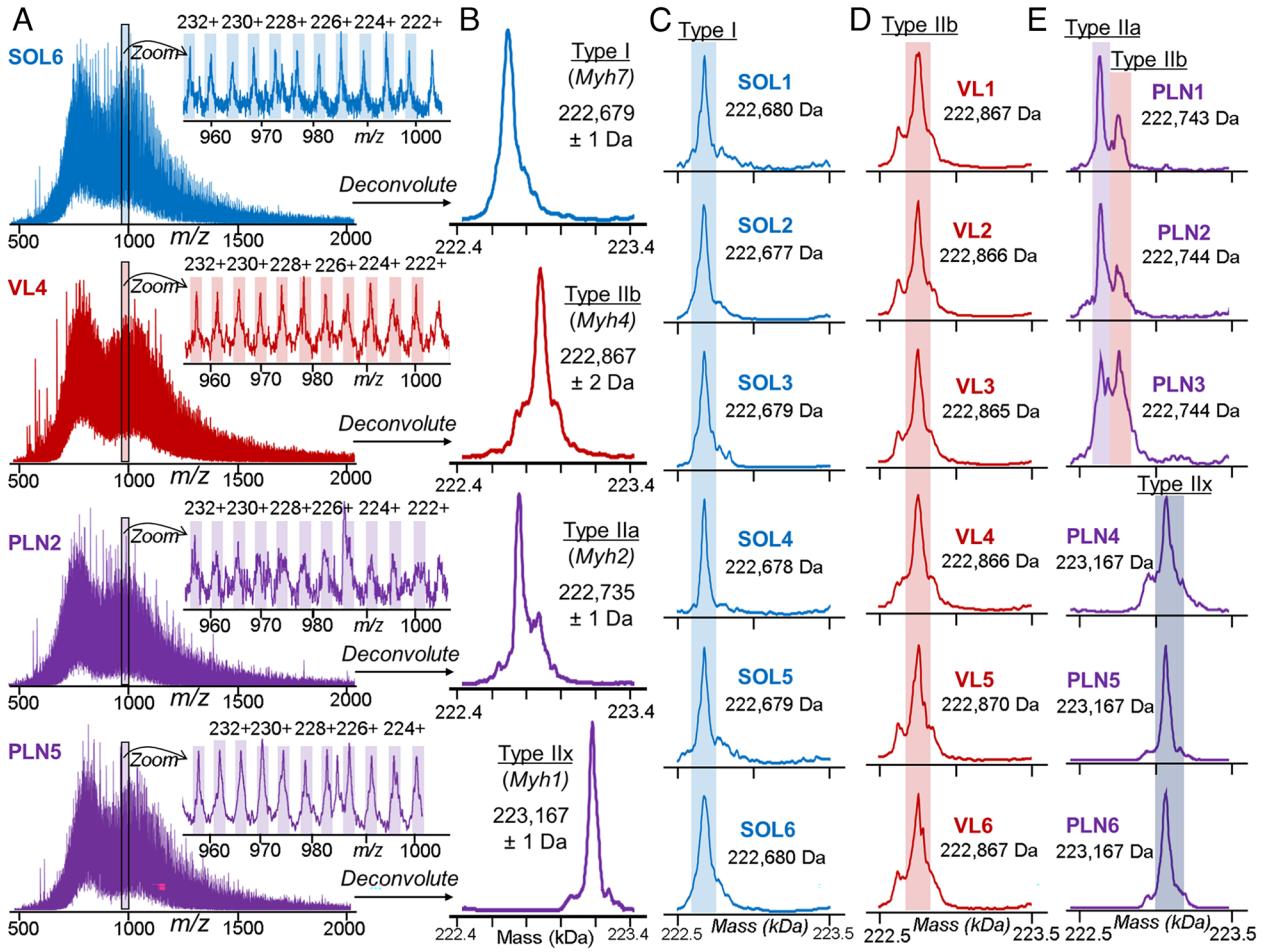

Fig. 3. Myosin heavy chain (MyHC) isoforms detected from SMFs. (*A*) Representative MS1 scans of MyHC isoforms from SMFs obtained from SOL, VL, and PLN muscles. Zoom-in of 960 to 1,000 *m/z* region shows the highly charged ions characteristic of large proteins with high accuracy mass measurements (1 to 2 Da). (*B*) Low-resolution maximum entropy deconvolution of representative VL, PLN, and SOL SMFs reveals four distinct masses from the MS1 spectra presumably corresponding to type I, type IIa, type IIb, and type IIx MyHC isoforms. (*C*–*E*) Deconvoluted mass spectra for all SMFs from SOL, VL, and PLN muscles (n = 6 fibers per muscle). (*C*) SOL, type I; (*D*) VL, type IIb; and (*E*) PLN, type IIa, type IIb, and Type IIx. SOL, VL, and PLN muscles (n = 6 fibers per muscle). (*C*) SOL, type I; (*D*) VL, type IIa and type IIb; and (*E*) PLN, type IIa, type IIb, and type IIx.

the VL and PLN muscles contained several highly abundant fsTnT3 proteoforms, including monophosphorylated fsTnT3 (pfsTnT3), fsTnT3, and pfsTnT3 with the loss of phosphate (ΔH_3PO_4) proteoforms (Fig. 4*A*). The fsTnT3 total level of phosphorylation (Ptotal) values (Fig. 4*C* and *SI Appendix*, Fig. S12) was highly consistent across all of the SMF samples obtained from VL and PLN muscles, as well as the relative expression of fsTnT3 (Fig. 4*D*). Appreciably, when we zoomed in on the VL and PLN deconvoluted spectra 20-fold, there were several low-abundance and fiber-specific fsTnT isoforms and proteoforms revealed (Fig. 4*B*). We detected fsTnT4 proteoforms from all VL and PLN, but the Ptotal values were highly variable, particularly for the SMFs obtained from PLN muscle (Fig. 4*C*). Besides heterogeneity in fsTnT4 proteoform expression, we found that some of the SMFs obtained from PLN contained unique fsTnT isoforms, including fsTnT9 and fsTnT10. These observations were further exemplified by normalizing all of the low-abundance proteoforms to their highest intensity peak (*SI Appendix*, Fig. S25).

Other examples of myofilament isoforms together with PTMs include thin filament protein α- and β-Tpm isoforms (Fig. 2*D* and *SI Appendix*, Fig. S15), as well as fast- and slow-twitch MLC isoforms (*SI Appendix*, Figs. S16–S19). Importantly, the top–down approach enabled the direct differentiation of two Tpm isoforms, β-Tpm and β′-Tpm (53), with only 26 Da mass difference (Fig. 2*D* and *SI Appendix*, Fig. S15). As illustrated, top–down proteomics is uniquely suited to provide a "bird's eye view" of the myriad isoforms together with PTMs from single muscle cells.

Top–Down Proteomics Captures Single-Cell Heterogeneity in SMFs at the Proteoform Level. The top–down proteomics of SMFs here clearly reveal single-cell proteoform heterogeneity (Fig. 5 and *SI Appendix*, Fig. S15, S18, S19, and S24). For example, the fast and slow isoforms of MLC-2 (MLC-2F and MLC-2S) displayed a remarkable amount of proteoform heterogeneity across the fiber samples (Fig. 5). The top–down approach detected MLC-2F and pMLC-2F in the SMFs obtained from VL and PLN, whereas MLC-2S and pMLC-2S were detected in SMFs obtained from SOL (Fig. 5 *A*–*C* and *SI Appendix*, Fig. S17). MLC-2S and MLC-2F have distinct functions yet share similar sequences differing by only 88.9 Da; even so, the isoforms can be separated by our online LC-MS/MS method chromatographically (Fig. 5*D*). The deconvoluted spectra clearly show that there are fiber-to-fiber differences in the Ptotal of MLC-2F and MLC-2S (Fig. 5*E*). On average, the MLC-2 Ptotal values for SMFs obtained from VL, PLN, and SOL were 0.15 ± 0.02, 0.34 ± 0.07, and 0.17 ± 0.08, respectively (Fig. 5*E*). The SMFs obtained from PLN and SOL displayed a remarkable degree of phosphorylation heterogeneity with MLC-2 Ptotal values ranging from 0.21 to 0.39 for SMFs obtained from PLN and 0.05 to 0.29 for SMFs obtained from SOL. Previously, top–down proteomics of bulk skeletal muscle

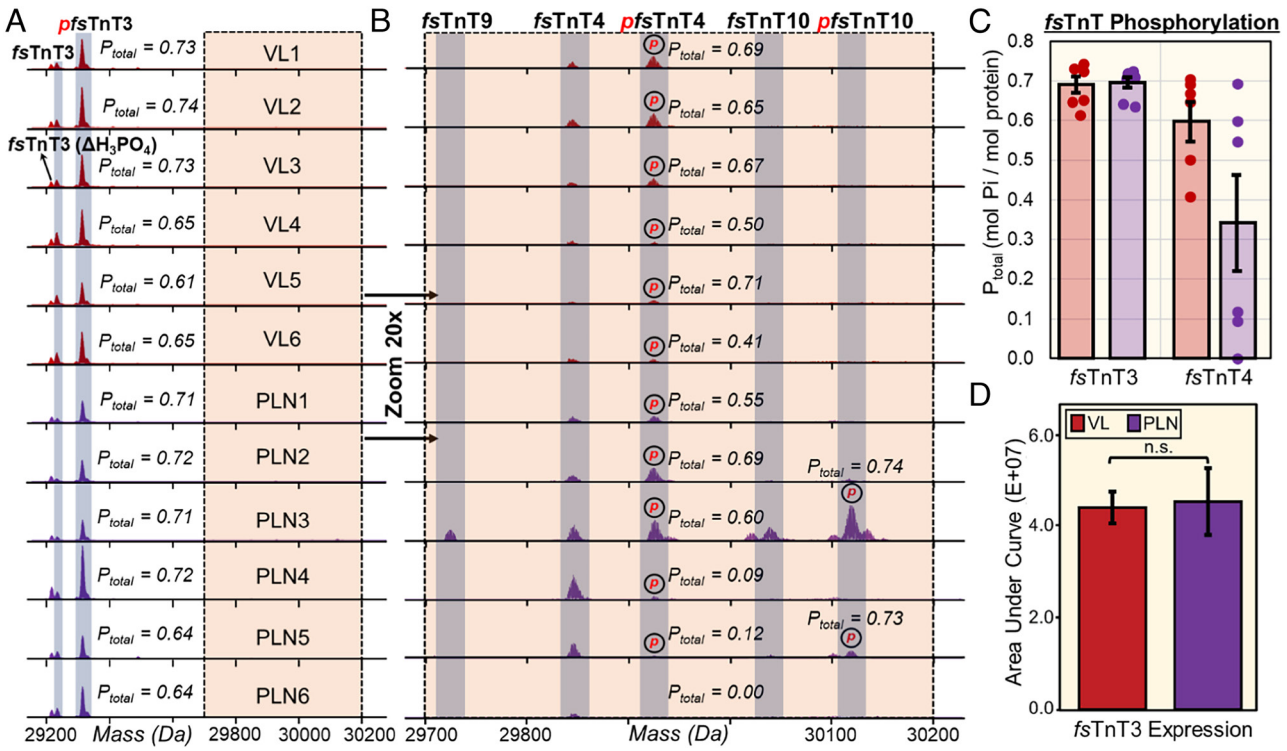

Fig. 4. Altered phosphorylation of low-abundance fast skeletal troponin T (fsTnT) isoforms and proteoforms from fiber-to-fiber. (*A*) Representative deconvoluted mass spectra of fast skeletal troponin 3 (*fs*TnT3) from SMFs isolated from fast-twitch VL (red) and PLN (purple) muscles. All the spectra are normalized to 1,000,000 intensity units. Mono-phosphorylation is denoted with red "p"; "ΔH_3PO_4" indicates phosphate loss from pfsTnT3. (*B*) 20× magnitude zoom-in on 29,700 Da to 30,200 Da region of spectra reveals several low-abundance fsTnT isoforms (*fs*TnT4, *fs*TnT9, *fs*TnT10) in SMFs from fast-twitch VL (red) and PLN (purple) muscles. All the spectra are normalized to 50,000 intensity units. Monophosphorylation is denoted with red "p". (*C*) Total phosphorylation (P_{tot}) calculated as mol Pi/mol protein for *fs*TnT3 and *fs*TnT4 from each fiber (n = 6). (*D*) Extracted ion chromatograms (EICs; top 5 most abundant ions) of *fs*TnT3 were made and the area under the curve was integrated to calculate the *fs*TnT3 expression. Groups were considered statistically different at $P < 0.05$; "n.s." indicates statistically not significant by the paired Student *t* test.

tissue samples reports an ensemble measurement of hundreds of fibers (37, 38, 56), which averages all the differences among diverse single muscle cells.

Moreover, fiber-to-fiber differences were also present in other sarcomeric proteoforms. For instance, both SMFs obtained from VL and PLN expressed the slow-twitch associated isoform, β-Tpm; however, SMFs obtained from PLN had a higher abundance (*SI Appendix*, Fig. S15). Furthermore, we found the MLC-3F was in all the SMFs obtained from SOL but at a much lower abundance than found in the SMFs obtained from PLN (*SI Appendix*, Fig. S18). Additionally, one SMF obtained from PLN (PLN1) expressed slow-twitch MLC-1V, whereas none of the other SMFs obtained from PLN contained this proteoform (*SI Appendix*, Fig. S19). Lastly, the ratio of MLC-1F to MLC-3F isoform expression, thick filament proteins involved in the structural stability of MyHC, was found to be highly heterogeneous across the SMFs obtained from VL and PLN muscles (*SI Appendix*, Fig. S26), implicating the structural role of isoform abundance in myosin stability. The myofilament proteoform heterogeneity observed in our samples is uniquely captured by performing top–down proteomics at the single fiber level.

Effective Sequence Characterization by Top–Down Online LC-MS/MS in SMF. The capability of the top–down online LC-MS/MS for proteoform identification and characterization at the single muscle level was demonstrated using the myriad MLC isoforms (Fig. 6) found in single muscle cells. MLCs are associated with MyHCs and essential for physiological speeds of shortening skeletal muscle (41). SMFs from SOL contained two MLC-1 isoforms (MLC-1S and MLC-1V) and SMFs from PLN and VL contained one MLC-1 isoform (MLC-1F) (*SI Appendix*, Figs. S16 and S19). Alignment of the sequences of MLC-1F, MLC-1S, and MLC-1V revealed the high homology of these isoforms (Fig. 6*A*). The LC-MS/MS method effectively separates all three of these isoforms from one another (Fig. 6*B*) and measures MLC-1 isoform spectra with high mass accuracy and isotopic resolution (Fig. 6*C*) prior to CAD MS/MS, which generates several unique *b* and *y* ions that are characteristic of the MLC-1 isoform sequences (Fig. 6*D*). Top–down proteomics enabled the characterization of each MLC-1 isoforms as well as their N-terminal modifications, such as acetylation and dimethylation (Fig. 6*E*). The percentage of interresidue bond cleavages of MLC-1F, MLC-1S, and MLC-1V were 14.4%, 19.5%, and 25.3%, respectively. In addition to MLC-1, we characterized the fast- and slow-twitch isoforms of MLC-2 (*SI Appendix*, Fig. S27). SMFs from SOL contained MLC-2S (20.1% bond cleavage) whereas SMFs from VL and PLN contained MLC-2F (25.2% bond cleavage) (*SI Appendix*, Fig. S17). Lastly, we were able to characterize MLC-3F (17.6% bond cleavage), which was found in every SMF sample (*SI Appendix*, Fig. S18), with our online LC-MS/MS method (*SI Appendix*, Fig. S28). Significantly, our top–down proteomics approach is able to characterize the entire MLC family including a large variety of isoforms.

Next, we have showed characterization of several key sarcomeric proteins including α-sActin, TnI, TnT, Tpm, and Cypher isoforms (*SI Appendix*, Fig. S29–S33). Importantly, we have demonstrated the capability of this top–down proteomics platform for characterization of large proteoform (>30 kDa) from single muscle cells at the chromatographic time scale. For example, online LC-MS/MS was successfully performed on a critical thin filament protein, α-sActin (Calc'd: 41,845.83 Da, Expt'l: 41,846.03 Da, 8.8%

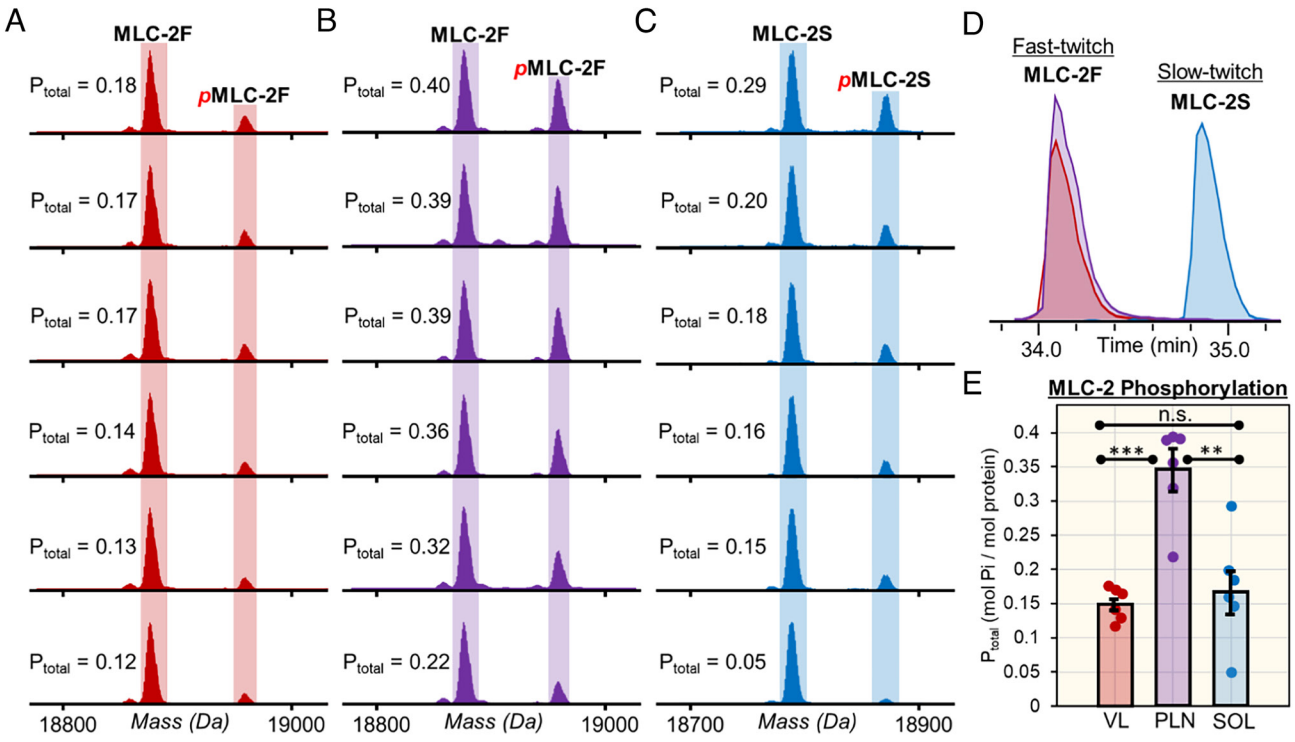

Fig. 5. Proteoform heterogeneity in myosin light chain 2 (MLC-2) isoforms across different SMFs. (*A*) EICs (top 5 most abundant ions) for fast and slow isoforms of myosin light chain 2 (MLC-2F and MLC-2S, respectively) obtained from SMFs from VL, PLN, and SOL muscles. (*B* and *C*) Deconvoluted mass spectra of MLC-2F from SMFs from fast-twitch VL and PLN muscles. (*D*) Deconvoluted mass spectra of MLC-2S from SMFs from slow-twitch SOL muscles. (*E*) Total phosphorylation (P_{tot}) calculated by mol Pi/mol protein for MLC-2F and MLC-2S for SMFs from VL, PLN, and SOL muscles (n = 6 fibers per muscle). Groups were considered statistically different at $P < 0.05$; "n.s." indicates statistically not significant by the paired Student *t* test.

bond cleavage) resulting in multiple high-quality *b* and y ions for sequence characterization (*SI Appendix*, Fig. S29). Top–down MS characterized highly similar isoforms such as the fast- and slow-twitch isoforms of TnI (10.8% and 13.3% bond cleavage for ssTnI and fsTnI), TnT, and Tpm (*SI Appendix*, Figs. S30–S32). Accurate mass measurements were obtained for the ssTnT (Calc'd: 31,187.07 Da, Expt'l: 31,187.18 Da, 5.8% bond cleavage) and fsTnT3 (Calc'd: 29,299.33 Da, Expt'l: 29,299.45 Da, 7.7% bond cleavage) together with online MS/MS for sequence characterization; similarly, LC-MS/MS data were obtained for β-Tpm (Calc'd: 32,858.56 Da, Expt'l: 32,858.68 Da, 5.6% bond cleavage) and α-Tpm (Calc'd: 32,702.68 Da, Expt'l: 32702.87 Da, 6.0% bond cleavage). Moreover, our online LC-MS/MS method allowed for characterization of the Z-disk proteins Cypher2s (Calc'd: 31,319.00 Da, Expt'l: 31,319.10 Da, 9.4% bond cleavage) and Cypher4s (Calc'd: 30879.76 Da, Expt'l: 30879.96 Da, 10.4% bond cleavage), which share highly similar sequences (98.3% sequence homology) (*SI Appendix*, Fig. S33).

The bond cleavages of the proteoforms characterized in this study range from 5.8 to 25.2% (*SI Appendix*, Table S2), which reflects the current limitations of performing top–down MS/MS on a chromatographic time scale. In particular, it is difficult to achieve comprehensive fragmentation in the online LC-MS/MS mode especially for large proteoforms >30 kDa (61). Typically, comprehensive fragmentation was achieved in the offline MS/MS mode (62, 63). Despite the relatively low number of bond cleavages, we have achieved several complementary *b/y* fragment pairs that cover the full sequence (Fig. 6 and *SI Appendix*, Figs. S27, S28, and S30–S32), representing the unique advantage of top–down MS as compared to bottom–up. Here, with high-accuracy MS1 measurements, highly reproducible LC separations, the complementary *b/y* fragment ion pairs, and annotations from previous studies (38, 56, 57), the identifications of proteoforms are confident. To increase the number of bond cleavages, alternative fragmentation techniques such as electron-based dissociation (i.e., electron capture dissociate/electron transfer dissociation) or absorption of high-energy photons (i.e. UVPD) could be implemented (63).

Additionally, we searched the MS/MS data obtained from a representative VL, PLN, and SOL fiber and performed Gene Ontology analysis (*SI Appendix*). The results showcase many processes related to muscle contraction and mitochondrial functions found across all fiber samples, as well as nuclear proteins, such as histones, as well as mitochondrial proteins (*SI Appendix*, Fig. S34 and Datasets S1–S3). Top–down proteomics provides a robust platform for the characterization of SMF proteins.

Discussion

In this study, we have developed a high-sensitivity top–down proteomics method including a one-pot sample processing to reduce protein losses, followed by capillary LC-MS/MS using low flow capillary RPLC and the MnESI source with detection limit as low as 0.5 ng total loading of a standard protein (Fig. 1 and *SI Appendix*, Fig. S7). Our approach is partly motivated by previous bottom–up single-cell proteomics studies which have shared a common theme of minimizing sample handling to avoid adsorptive losses, incorporating highly sensitive protein separation methods, and exploiting ultrahigh-resolution MS instrumentations (8–11, 21, 22, 44–46). We have shown the effective separation of large proteins and characterization of proteoforms by online LC-MS/MS with high reproducibility (Fig. 2 and *SI Appendix*, Figs. S8–S11). On average, SMF (multinucleated single muscle cell) extracts contained ~180 ng total protein per fiber (*SI Appendix*, Table S3). In comparison, in the early single-cell bottom–up proteomics studies, large cell types such as blastocytes were used where there can be

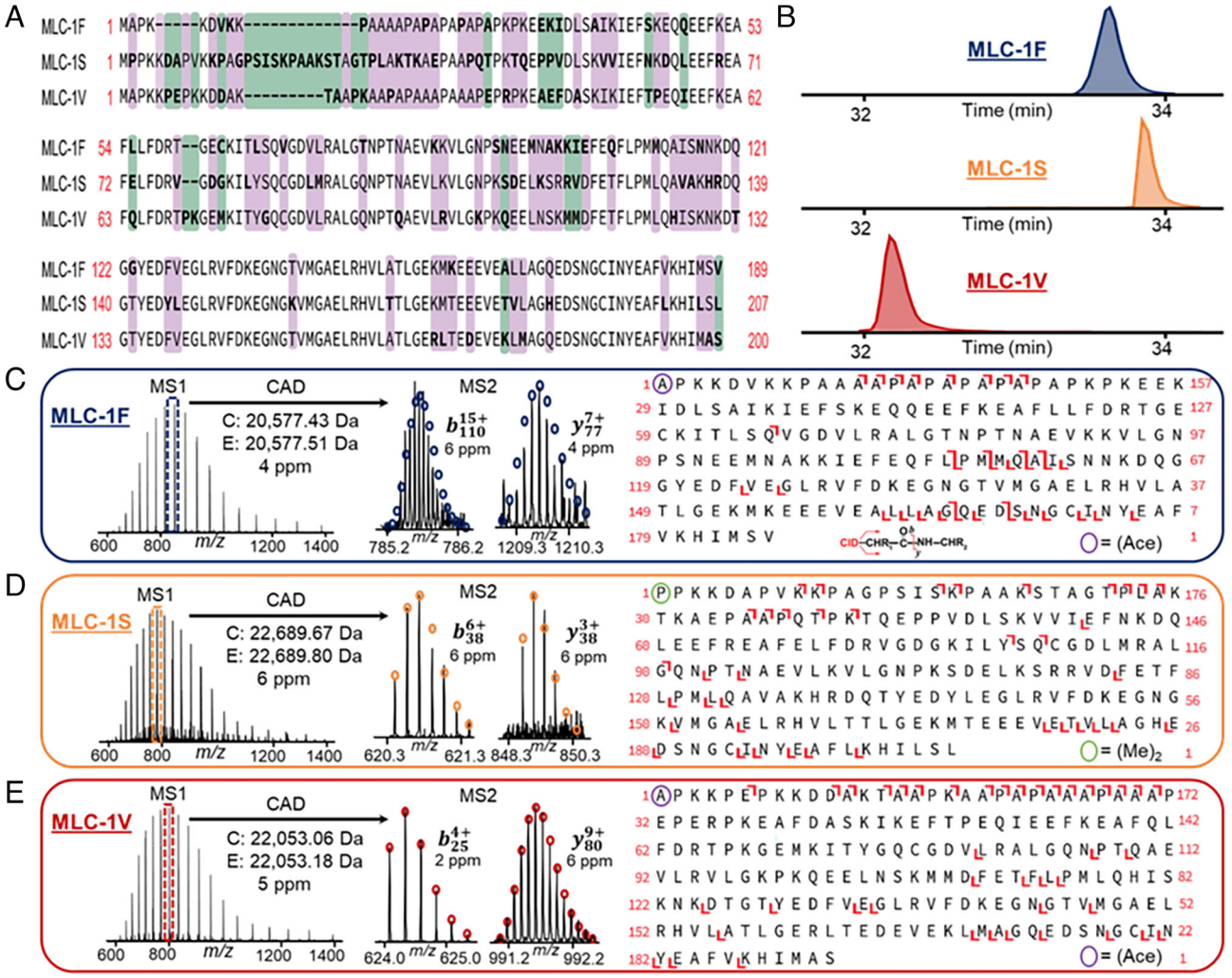

Fig. 6. Top–down MS characterization of myosin light chain 1 (MLC-1) isoforms from SMFs. (*A*) Sequence alignment of the fast, slow, and ventricular isoforms of myosin light chain 1 (MLC-1F, MLC-1S, and MLC-1V, respectively). Purple indicates no residues shared in common between the isoforms, green indicates at least one residue is shared between the isoforms, and no color indicates sequence homology between the isoforms. (*B*) Representative extracted ion chromatograms (EICs) for MLC-1 isoforms. The EICs of MLC-1F, as well as MLC-1S and MLC-1V, are from SMFs from VL and SOL muscles, respectively. (*C–E*) Online LC–MS/MS of MLC-1 isoforms. Precursor ions were selected for online collisionally activated dissociation resulting in signature *b* and *y* ions that are characteristic of MLC-1F (*C*), MLC-1S (*D*), and MLC-1V (*E*). The bond cleavages of MLC-1F (11 *b* ions, 23 *y* ions, *P* value: 5.1×10^{-21}), MLC-1S (18 *b* ions, 24 *y* ions, *P* value: 4.5×10^{-33}), and MLC-1V (18 *b* ions, 35 *y* ions, *P* value: 2.5×10^{-37}) were 14.4%, 19.5%, and 25.3%, respectively. "(Ace)-" denotes N^{α}-acetylation, "$(Me)_2$-" denotes N^{α}-dimethylation. Circles represent the theoretical isotopic abundance distribution corresponding to the assigned mass and based on the averagine model.

up to 10 μg of total protein in the cell (7, 64). The most recent state-of-the-art technologies are able to analyze as little as 0.2 ng of tryptic peptides from smaller cell types such as HeLa or acute myeloid leukemia cell lysates (10, 12, 21, 44, 45). It is known that top–down proteomics is inherently less sensitive compared to bottom–up proteomics, due to the exponential decay in S/N as a function of molecular weight and the adsorptive nature of proteins (17, 28).

We have applied the top–down proteomics method to skeletal muscle SMFs (multinucleated single cells) because 1) skeletal muscle fibers are remarkably heterogeneous at the molecular level and, thus, provide a "proteoform-rich" proteome for top–down study (32–35), 2) the fusion of skeletal muscle precursor cells, known as myoblasts, during development produces relatively large cells, approximately equivalent to 1 to 2×10^3 mononucleated single cells, with greater protein content than most other somatic cell types (34, 35), and 3) the structural and functional diversity of skeletal muscle is reflected in its basic contractile unit, the sarcomere which contains a variety of isoforms from multigene families together with PTMs making it an ideal system to leverage the unique capability provided by our top–down approach (37, 38, 41, 56, 65). As demonstrated, the single-cell top–down proteomics analysis provides a bird's eye view of protein isoforms together with PTMs in conjunction with functional measurements and captures single muscle cell heterogeneity at the proteoform level (Fig. 2 *B–D*).

Skeletal muscle fibers have typically been classified as either slow/type I and fast/type II based on their physical, morphological, histological, and biochemical characteristics (42). Previous studies have shown that muscle fiber classification, and thereby function, is largely defined by which MyHC isoform expressed within the fiber (i.e., type I, IIa, IIb, or IIx) (34, 35, 40–43). In rat, there are four MyHC isoforms expressed; fibers may express only one of these isoforms ("pure") or may express more than one isoform ("hybrid") (35, 54). Variances in force generation have been linked with altered expression of MyHC isoforms demonstrating that SMF contractile performance and proteoform landscape forms a continuum (34, 41). We report the detection of MyHC isoforms (~223 kDa) in SMF with single-cell resolution enabled by the highly reproducible intact mass analysis (Fig. 3 and *SI Appendix*,

Fig. S20). Although several MyHC isoforms have been studied from SMFs using bottom–up proteomics (32, 33, 66) or middle–down MS (67), intact MyHC isoforms have never been detected at the whole protein level in SMFs. To connect the MyHC proteomics data to the functional measurements, the fast-twitch fibers (VL and PLN) had a greater maximum shortening velocity than slow-twitch fibers (SOL) (Fig. 2*B*), which is strongly correlated with MyHC isoform expressions (Fig. 3). Importantly, there was greater variance in the VL and PLN fiber functional measurements because multiple MyHC isoforms may be expressed in these fibers.

The intact mass measurements for MyHC isoforms (~223 kDa) were highly consistent and reproducible across similar fiber types with SDs of 1 to 2 Da. We note the experimentally determined masses for these isoforms did not exactly match those calculated based on sequences in the UniProt database, implying potential sequence modifications (34, 35, 40–43). Online LC-MS/MS fragmentation was attempted but unfortunately did not yield fragmentation ions sufficient for confirmation of MyHC sequence, PTM, or isoform assignment. With such large proteins (~223 kDa), MS/MS is extremely challenging due to the formation of compact gas phase structures that are refractory to fragmentation (17, 68). This difficulty is compounded when trying to achieve such fragmentation on a chromatographic timescale from single cells. Thus, MyHC isoform assignments were based on the consistent intact mass analysis (MS1 mode), as well as the well-documented expression patterns of these isoforms in the specific muscle types (14–17). Rat SOL and VL muscles are composed almost exclusively of fibers expressing type I and type IIB MyHC, respectively (35, 41, 54). Thus, the lone MyHC isoforms detected in fibers from SOL and VL were assigned as type I (*Myh7*) and type IIB (*Myh4*) MyHC, respectively. PLN is a muscle composed of all four fiber types (type I, IIX, IIB, and IIA), yet, type IIA (~44%) and type IIX/IIB (~16 to 20%, each) are predominant (35, 41, 54). We detected three isoforms with masses of 222,735 ± 1 Da, 222,867 ± 2 Da, and 223,167 ± 1 Da in fibers from PLN muscles. The 222,858 Da isoform had similar mass to the sole isoform detected in VL and was therefore assigned type IIB. Assignment of the last two isoforms was based on calculated molecular weights for these isoforms. Based on mass calculation, type IIX is larger than type IIA MyHC. Thus, the isoforms with masses 223,168 Da and 222,735 Da were assigned as type IIX and type IIA, respectively.

Despite the lack of MS/MS, it is quite certain in the assignment of proteins with intact mass of ~223 kDa as MyHC because MyHC has a mass of ~223 kDa and is most abundant protein representing a substantial portion (>40%) of the total protein in the skeletal muscles (32, 69–71). Given the limited sample amount in a single muscle cell, only the most abundant proteins could be detected in the 1DLC-MS/MS study here. Moreover, since it is well-documented on the specific MyHC isoforms expressed in rat SOL, VL, and PLN muscles, we are quite confident about the assignment of the MyHC isoforms in this study. MyHC is also known to be heavily modified (34, 35, 40–43) so it is conceivable that mass discrepancy between the experimentally determined masses for MyHC isoforms and those calculated based on isoform sequences in the database likely is a consequence of PTMs (34, 35, 40–43). Nevertheless, future studies with technology advances will be necessary to unambiguously determine the identity of these isoforms and characterized the full sequences and PTMs.

Our top–down proteomics data illustrate the heterogeneous nature of SMFs and captures single muscle cell heterogeneity at the proteoform level. Heterogeneity in MyHC isoform expression was observed, particularly for the fast-twitch PLN muscle where multiple MyHC isoforms were detected (types IIa, IIb, and IIx) (Fig. 3). Additionally, we found that the expression of MLC-1F and MLC-3F varied from fast-twitch fiber-to-fiber (*SI Appendix*, Fig. S26). This heterogeneity in thick filament proteins is noteworthy because both MyHC and MLC isoforms determine the maximum shortening velocity of SMFs (41, 65). In fact, the functional data demonstrated extensive heterogeneity in shortening velocity, particularly in the fast-twitch fibers, which can be related to the proteomics heterogeneity (Fig. 2*B*). In addition, low abundance fsTnT isoforms together with PTMs as well as MLC-2 proteoforms demonstrated fiber-to-fiber differences from the bird's eye view provided by top–down proteomics (Figs. 4 and 5). Besides fsTnT and MLC-2 proteoform heterogeneity, we demonstrate that several other proteins, including both myofilament and Z-disk proteins (*SI Appendix*, Figs. S15–S30), display varying degrees of proteoform heterogeneity as uniquely enabled by our top–down approach. Our results clearly demonstrate that there are fiber-to-fiber differences in proteoforms, whereas in bulk tissue analysis the results represent an ensemble measurement of hundreds of fibers in addition to other components found in skeletal muscle.

Notably, in this study, we have been able to detect large proteoforms (>200 kDa) from a single muscle cell. Although a previous study by Zhu et al. has used the nanoPOTs platform for the extraction and sampling of intact proteins from low numbers of HeLa cells, only small proteoforms (<15 kDa) were detected (72). Recently, a capillary electrophoresis coupled to MS/MS (CE-MS/MS) approach has been used for the analysis of proteins from single cells; nonetheless, the detected proteoforms were almost exclusively below 15 kDa with no proteins greater than 30 kDa (73). Top–down MS analysis of large proteins is challenging due to an exponential decay in S/N with increasing molecular weight (MW) as well as coelution with low MW and high-abundance proteins in a mixture (17, 28, 74). Typically, size-based fractionation prior to MS analysis is essential for the detection of large proteins. We have previously developed a top–down 2D-LC method including serial size exclusion chromatography to enrich high MW proteins with reverse phase chromatography and high-resolution top–down MS for the detection of large proteins (74, 75). We have detected large proteins (>30 kDa) using a simple 1D capillary RPLC-MS from single muscle cells and reveal heterogeneity at the level of proteoforms.

The SMF proteome has a large dynamic range and is dominated by the contractile machinery of muscle, which accounts for ~54% of the total protein mass of skeletal muscle (32, 69, 70) (Dataset S4). By taking a top–down 1D LC-MS/MS approach with only 100 ng total loading, the highly abundant proteins of the contractile machinery from the skeletal muscle were predominantly detected in this study. In fact, this is a common limitation for top–down proteomics that mainly the highly abundant proteins could be detected in the 1DLC-MS/MS approach regardless the amount of samples injected. The previous top–down proteomics studies on rat muscle proteome injected more sample (300 to 1,000 ng) but still identified the most abundant (and lower molecular weight) proteins due to the complexity and high dynamic range of the muscle proteome (38, 56, 57). Conceivably, there is a limited capacity in separation of intact proteins in the 1DLC mode. To achieve the deep proteome coverage, multiple dimensions of separation are required as demonstrated in previous studies (74, 76, 77).

Conclusion

Here, we developed a top–down proteomics method that combined one-pot sample preparation and highly sensitive capillary LC-MS/MS. We have shown the effective separation of proteins and characterization of large proteoforms on the chromatographic time scale

with reproducibility. This method was applied to SMFs (multinucleated single cells) with remarkable functional and proteomics heterogeneity. We have integrated functional measurements and top–down proteomic analysis of SMFs and captured single muscle cell heterogeneity at the proteoform level. Using SMFs obtained from three functionally distinct muscles, we found fiber-to-fiber heterogeneity among the sarcomeric proteoforms and striking heterogeneity in large proteoforms (>200 kDa). Moreover, this study underlined the unique capability of top–down proteomics in simultaneous characterization of PTMs together with a variety of isoforms from multigene families in the sarcomere. Impressively, we detected multiple distinct MyHC isoforms (>220 kDa) from the SMFs with high reproducibility and robustness which enables the classification of fiber type at the single-cell level. This study reports single-cell resolution of large proteoforms, highlighting the potential of single-cell top–down proteomics in advancing our understanding of phenotypic heterogeneity and functional diversity among individual cells toward precision medicine. We envision that this top–down proteomics method developed can be extended to other applications requiring high sensitivity.

Materials and Methods

Detailed materials and methods are outlined in *SI Appendix, Methods*.

Reagents and Chemicals. All reagents were purchased from Sigma-Aldrich, Inc. unless otherwise noted. High performance liquid chromatography (HPLC)-grade water and acetonitrile were purchased from Fisher Scientific. Hexafluro-2-propanol (HFIP) was purchased from Alfa Aesar. Protease inhibitor cocktail was purchased from Thermo Fisher Scientific. Amicon, 0.5-mL cellulose centrifugal filters with a 10 kDa molecular weight cutoff (MWCF) were purchased from MilliporeSigma.

Animals. Male Fischer 344 Norway F1 hybrid (F344BN) rats (aged 6 mo) were obtained from the University of Wisconsin-Madison School of Veterinary Medicine. The rats were given access to food and water ad libitum and individually housed in clear plastic cages on a 12 h/12 h light/dark cycle. All procedures involving animals were carried out following the recommendations in the Guide for the Care and Use of Laboratory Animals published by the NIH and were approved by the Institutional Animal Care and Use Committee of the University of Wisconsin-Madison.

Relaxation Solution. The relaxation solution used for the isolation of SMFs was similar to that described previously (78) (2 mM EGTA, 100 mM KCl, 10 mM imidazole, 4 mM ATP, 5 mM $MgCl_2$) with minor modifications to accommodate top–down proteomics studies, including the addition of antioxidants [5 mM tris(2-carboxyethyl)phosphine), as well as protease (2 mM phenylmethylsulfonyl fluoride) and phosphatase inhibitors (2 mM sodium vanadate and 10 mM β-glycerophosphate)].

Isolation of SMFs. Bundles of ~50 fibers were dissected from the VL, PLN, and SOL muscles and placed in a petri dish containing relaxation solution under a confocal microscope. For each experiment, an individual fiber (~1.5 to 2.5 mm in length) was pulled from the end of the bundle using fine-point tweezers and placed in a labeled low-protein binding microcentrifuge tube. Before it was placed in the tube, each isolated fiber was examined under the microscope and determined to be a single fiber that was intact (undamaged) (*SI Appendix*, Fig. S35). The location of the fiber in the tube was outlined with marker prior to freezing at −80 °C until the day of the experiment. In total, there were 10 fibers per tissue type used for mechanical measurements and 6 fibers per tissue type used for top–down proteomics analysis. All fibers were isolated from the muscles of a single rat.

Shortening Velocity Measurements of SMFs Obtained from VL, PLN, and SOL. The experimental technique for performing contractile measurements on skeletal muscle fibers has been described previously (56, 78). Briefly, the fiber segments were attached between a capacitance-gauge transducer (Model 403, sensitivity of 20 mV/mg and resonant frequency 600 Hz; Aurora Scientific) and a DC torque motor (Model 308; Aurora Scientific). Length changes during contractile measurements were introduced at one end of the preparation driven by voltage commands from a PC via a 16-bit D/A converter.

The velocity of unloaded shortening (Vo) was measured using the slack-test method described previously by Edman (79). With the fiber length at L, the fiber was transferred to pCa 4.5 solution and steady force was allowed to develop. The fiber was then quickly released to a shorter length, resulting in an immediate tension decline, followed by a gradual recovery of tension. The time interval before tension was redeveloped was plotted against the length of release for several different releases, and the slope of this line was calculated as the velocity of unloaded shortening. All velocity values were converted from millimeters per second to fiber lengths per second by dividing by L.

Extraction of sarcomeric proteins from SMFs. SMFs from the VL, PLN, and SOL muscles were thawed on ice the day of LC-MS/MS experiments. To minimize artificial protein modifications and oxidation, protein extraction was performed at 4 °C. Samples were washed with 40 μL of 150 mM ammonium acetate to remove any remaining relaxation buffer by pipetting the solution over the fiber such that the fiber remained on the wall of the tube. Samples were briefly centrifuged (1,100 × *g*, 1 min, 4 °C) and the solution was removed. The fiber was suspended in 20 μL of protein extraction solution (25% HFIP, 10 mM L-methionine, 1× HALT protease inhibitor cocktail, and 1× phosphatase inhibitor cocktail A, pH 7.5). The samples were incubated for 15 min on ice followed by addition of 20 μL of mobile phase A (MPA; 0.1% formic acid (FA) in water) and a freeze–thaw cycle (incubation at −80 °C for 5 min followed by incubation for 1 min at 37 °C). The freeze–thaw cycle was repeated two additional times for a total of three cycles with mixing by gentle agitation between cycles. Samples were centrifuged (21,000 × *g*, 15 min, 4 °C). The fiber extracts were desalted prior to LC-MS/MS analysis using a 10 kDa MWCF and buffer exchanged into MPA concentrating to a final volume of 35 μL. Then, 10 μL of the desalted extracts was used for a Bradford protein assay to estimate the protein concentration. The remaining 25 μL was transferred to a HPLC vial.

Online LC-MS/MS Analysis of SMF Extracts. LC-MS/MS analysis was carried out using a NanoAcquity ultra-high pressure LC system (Waters) coupled to a high-resolution maXis II quadrupole time-of-flight mass spectrometer (Bruker Daltonics). A microflow-nanospray electrospray ionization source (MnESI) equipped with multinozzle M3 emitters (8 nozzles, 10 μm ID) provided by Newomics Inc. (Berkeley, California) was used for nanoESI. Approximately 100 ng of total protein extracted from SMF (per sample) was separated using a MAbPac™ Capillary Reversed-Phase HPLC Column (Thermo Scientific, 1,500 Å pore size, 4-μm particle size, 150-μm inner diameter). SMF proteoforms were eluted using a linear 60-min gradient going from 10 to 95% mobile phase B (0.1% FA in acetonitrile) at a constant flow rate of 2 μL/min. Data-dependent LC-MS/MS was performed on single fiber protein extracts. The three most intense ions in each mass spectrum were selected and fragmented by collisionally activated dissociation (CAD) with a scan rate of 2 Hz from 200 to 3,000 *m/z*. The isolation window for online AutoMS/MS CAD was 10 m/z. The collision DC bias was set from 20 to 40 eV for CAD with nitrogen as the collision gas. All singly charged ions were excluded for fragmentation and dynamic exclusion was enabled after the collection of 4 of the same precursor ions to prevent fragmentation of the same species. All identified proteins except MyHC were confirmed by MS/MS (*SI Appendix, Supplemental Note 6* and Table S2).

Data Analysis. DataAnalysis (Version 4.3; Bruker Daltonics) software was used to process and analyze the LC-MS data. The resolving power for maximum entropy deconvolution was set to 70,000 for proteins that were isotopically resolved and 10,000 for proteins that were not isotopically resolved. Monoisotopic masses are reported for all isotopically resolved MS and MS/MS data using the Sophisticated Numerical Annotation Procedure peak-picking algorithm. Most abundant masses are reported for all nonisotopically resolved MS data using the Sum Peak algorithm. The Ptotal for each phosphorylated protein was calculated using the following equation:

$$P_{total} = \frac{mol\ P_i}{mol\ protein}$$

The top five most abundant charge state ions of the sarcomeric proteins of interest were used to produce extracted ion chromatograms, and the area under the curve was integrated to determine the relative abundance for each protein. Tandem mass spectra were exported from DataAnalysis software and analyzed using MASH Explorer (Version 2.0) (80).

Data, Materials, and Software Availability. MS data have been deposited in MassIVE datasets (MSV000090967).

ACKNOWLEDGMENTS. We would like to acknowledge Kalina Reese for her contributions in designing the figures. Additionally, we would like to acknowledge Guillaume Tremintin (Bruker Daltonics) for his assistance in coupling the Newomics source with Bruker mass spectrometer. We would also like to thank Xuefei Sun and Shanhua Lin (Thermo Fisher Scientific) for providing the LC columns used in this study. Y.G. would like to acknowledge NIH R01 GM125085, R01 HL096971, R01 HL109810, GM117058, and S10 OD018475. W.G. would like to acknowledge the support from the NIH R01 HL148733, AHA 19TPA34830072, and USDA-NIFA AAI4641. J.A.M. acknowledges the support from the Training Program in Translational Cardiovascular Science, T32 HL007936-20 and T32 HL007936-21, for funding during the duration of this project. D.S.R. acknowledges the support from the American Heart Association Predoctoral Fellowship Grant No. 832615/David S. Roberts/2021. K.A.B. acknowledges the Vascular Surgery Research Training Program Grant T32 HL110853. E.A.C. acknowledges the support from the NIH Chemistry-Biology Interface Training Program T32 GM008505. Some aspects of the figures were created using BioRender.com.

Author affiliations: [a]Department of Chemistry, University of Wisconsin-Madison, Madison, WI 53706; [b]Department of Animal and Dairy Sciences, University of Wisconsin-Madison, Madison, WI 53706; [c]Department of Cell and Regenerative Biology, University of Wisconsin-Madison, Madison, WI 53705; [d]Department of Kinesiology, University of Wisconsin-Madison, Madison, WI 53706; [e]Human Proteomics Program, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, WI 53705; [f]Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI 53705; and [g]Newomics Inc., Berkeley, CA 94710

Author contributions: J.A.M., K.A.B., Y.J., G.M.D., and Y.G. designed research; J.A.M., D.S.R., E.A.C., L.E.E., E.J.L., Y.J., J.L., J.H., Y.Z., and G.M.D. performed research; D.W. contributed new reagents/analytic tools; J.A.M., K.A.B., Z.R.G., D.S.R., E.A.C., Z.G., Y.J., S.J.M., G.M.D., and Y.G. analyzed data; and J.A.M., K.A.B., Z.R.G., D.S.R., E.A.C., D.W., W.G., G.M.D., and Y.G. wrote the paper.

1. J. M. Perkel, Single-cell analysis enters the multiomics age. *Nature* **595**, 614–616 (2021).
2. R. Elmentaite, C. Domínguez Conde, L. Yang, S. A. Teichmann, Single-cell atlases: Shared and tissue-specific cell types across human organs. *Nat. Rev. Genet.* **23**, 395–410 (2022).
3. T. Stuart, R. Satija, Integrative single-cell analysis. *Nat. Rev. Genet.* **205**, 257–272 (2019).
4. N. Slavov, Unpicking the proteome in single cells. *Science* **367**, 512–513 (2020).
5. J. A. Alfaro *et al.*, The emerging landscape of single-molecule protein sequencing technologies. *Nat. Methods* **18**, 604–617 (2021).
6. V. Marx, A dream of single-cell proteomics. *Nat. Methods* **16**, 809–812 (2019).
7. R. M. Onjiko, S. A. Moody, P. Nemes, Single-cell mass spectrometry reveals small molecules that affect cell fates in the 16-cell embryo. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 6545–6550 (2015).
8. A. A. Petelski *et al.*, Multiplexed single-cell proteomics using SCoPE2. *Nat. Protoc.* **16**, 5398–5425 (2021).
9. Y. Zhu *et al.*, Nanodroplet processing platform for deep and quantitative proteome profiling of 10–100 mammalian cells. *Nat. Commun.* **9**, 882 (2018).
10. A.-D. Brunner *et al.*, Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation. *Mol. Syst. Biol.* **18**, e10798 (2022).
11. E. M. Schoof *et al.*, Quantitative single-cell proteomics as a tool to characterize cellular hierarchies. *Nat. Commun.* **12**, 3341 (2021).
12. R. T. Kelly, Single-cell proteomics: Progress and prospects. *Mol. Cell Proteomics* **19**, 1739–1748 (2020).
13. S. Aldridge, S. A. Teichmann, Single cell transcriptomics comes of age. *Nat. Commun.* **11**, 4307 (2020).
14. J. G. Camp, R. Platt, B. Treutlein, Mapping human cell phenotypes to genotypes with single-cell genomics. *Science* **365**, 1401–1405 (2019).
15. N. Slavov, Counting protein molecules for single-cell proteomics. *Cell* **185**, 232–234 (2022).
16. J. S. Brodbelt, Deciphering combinatorial post-translational modifications by top-down mass spectrometry. *Curr. Opin. Chem. Biol.* **70**, 102180 (2022).
17. J. A. Melby *et al.*, Novel strategies to address the challenges in top-down proteomics. *J. Am. Soc. Mass Spectrom.* **32**, 1278–1294 (2021).
18. R. Aebersold, M. Mann, Mass-spectrometric exploration of proteome structure and function. *Nature* **537**, 347–355 (2016).
19. Y. Zhang, B. R. Fonslow, B. Shan, M. C. Baek, J. R. Yates, Protein analysis by shotgun/bottom-up proteomics. *Chem. Rev.* **113**, 2343–2394 (2013).
20. F. Meissner, J. Geddes-McAlister, M. Mann, M. Bantscheff, The emerging role of mass spectrometry-based proteomics in drug discovery. *Nat. Rev. Drug Discov.* **21**, 637–654 (2022).
21. S. T. Gebreyesus *et al.*, Streamlined single-cell proteomics by an integrated microfluidic chip and data-independent acquisition mass spectrometry. *Nat. Commun.* **13**, 37 (2022).
22. C.-F. Tsai *et al.*, Surfactant-assisted one-pot sample preparation for label-free single-cell proteomics. *Commun. Biol.* **4**, 265 (2021).
23. D. L. Plubell *et al.*, Putting humpty dumpty back together again: What does protein quantification mean in bottom-up proteomics? *J. Proteome Res.* **21**, 891–898 (2022).
24. L. M. Smith, N. L. Kelleher, Proteoforms as the next proteomics currency: Identifying precise molecular forms of proteins can improve our understanding of function. *Science* **359**, 1106–1107 (2018).
25. L. M. Smith, N. L. Kelleher, Proteoform: A single term describing protein complexity. *Nat. Methods* **10**, 186–187 (2013).
26. R. Aebersold *et al.*, How many human proteoforms are there? *Nat. Chem. Biol.* **14**, 206–214 (2018).
27. L. M. Smith *et al.*, The human proteoform project: Defining the human proteome. *Sci. Adv.* **7**, 734 (2021).
28. P. D. Compton, L. Zamdborg, P. M. Thomas, N. L. Kelleher, On the scalability and requirements of whole protein mass spectrometry. *Anal. Chem.* **83**, 6868–6874 (2011).
29. B. Chen, K. A. Brown, Z. Lin, Y. Ge, Top-down proteomics: Ready for prime time? *Anal. Chem.* **90**, 110–127 (2018).
30. K. A. Brown, J. A. Melby, D. S. Roberts, Y. Ge, Top-down proteomics: Challenges, innovations, and applications in basic and clinical research. *Expert Rev. Proteomics* **17**, 719–733 (2020).
31. T. K. Toby, L. Fornelli, N. L. Kelleher, Progress in top-down proteomics and the analysis of proteoforms. *Annu. Rev. Anal. Chem.* **9**, 499–519 (2016).
32. M. Murgia *et al.*, Single muscle fiber proteomics reveals fiber-type-specific features of human muscle aging. *Cell Rep.* **19**, 2396–2409 (2017).
33. M. Murgia *et al.*, Signatures of muscle disuse in spaceflight and bed rest revealed by single muscle fiber proteomics. *PNAS Nexus* **1**, pgac086 (2022).
34. R. L. Moss, G. M. Diffee, M. L. Greaser, Contractile properties of skeletal muscle fibers in relation to myofibrillar protein isoforms. *Rev. Physiol. Biochem. Pharmacol.* **126**, 1–63 (1995).
35. S. Schiaffino, C. Reggiani, Fiber types in mammalian skeletal muscles. *Physiol. Rev.* **91**, 1447–1531 (2011).
36. P. M. Hwang, B. D. Sykes, Targeting the sarcomere to correct muscle function. *Nat. Rev. Drug Discov.* **145**, 313–328 (2015).
37. J. A. Melby *et al.*, Top-down proteomics reveals myofilament proteoform heterogeneity among various rat skeletal muscle tissues. *J. Proteome Res.* **19**, 446–454 (2020).
38. L. Wei *et al.*, Novel sarcopenia-related alterations in sarcomeric protein post-translational modifications (PTMs) in skeletal muscles identified by top-down proteomics. *Mol. Cell. Proteomics* **17**, 134–145 (2018).
39. Z. Wang *et al.*, The molecular basis for sarcomere organization in vertebrate skeletal muscle. *Cell* **184**, 2135–2150.e13 (2021).
40. R. L. Moss, R. J. Solaro, The enduring relationship between myosin enzymatic activity and the speed of muscle contraction. *J. Gen. Physiol.* **151**, 623–627 (2019).
41. S. Schiaffino, C. Reggiani, Myosin isoforms in mammalian skeletal muscle. *J. Appl. Physiol.* **77**, 493–501 (1994).
42. D. I. Resnicow, J. C. Deacon, H. M. Warrick, J. A. Spudich, L. A. Leinwand, Functional diversity among a family of human skeletal muscle myosin motors. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 1053–1058 (2010).
43. H. M. Warrick, J. A. Spudich, Myosin structure and function in cell motility. *Annu. Rev. Cell Biol.* **3**, 379–421 (2003), 10.1146/annurev.cb.03.110187.002115.
44. Y. Cong *et al.*, Ultrasensitive single-cell proteomics workflow identifies >1000 protein groups per mammalian cell. *Chem. Sci.* **12**, 1001–1006 (2021).
45. Y. Zhu *et al.*, Proteomic analysis of single mammalian cells enabled by microfluidic nanodroplet sample preparation and ultrasensitive nanoLC-MS. *Angew. Chemie* **130**, 12550–12554 (2018).
46. C. Ctortecka *et al.*, An automated workflow for multiplexed single-cell proteomics sample preparation at unprecedented sensitivity. bioRxiv [Preprint] (2022). https://doi.org/10.1101/2021.04.14.439828. (2021.04.14.439828).
47. W. Cai *et al.*, Temperature-sensitive sarcomeric protein post-translational modifications revealed by top-down proteomics. *J. Mol. Cell. Cardiol.* **122**, 11–22 (2018).
48. Z. Lin *et al.*, Comprehensive characterization of swine cardiac troponin T proteoforms by top-down mass spectrometry. *J. Am. Soc. Mass Spectrom.* **29**, 1284–1294 (2018).
49. Y. Peng *et al.*, Top-down proteomics reveals concerted reductions in myofilament and Z-disc protein phosphorylation after acute myocardial infarction. *Mol. Cell. Proteomics* **13**, 2752–2764 (2014).
50. A. C. Susa, Z. Xia, E. R. Williams, Native mass spectrometry from common buffers with salts that mimic the extracellular environment. *Angew. Chemie Int. Ed.* **56**, 7912–7915 (2017).
51. G. Huang, G. Li, R. G. Cooks, Induced nanoelectrospray ionization for matrix-tolerant and high-throughput mass spectrometry. *Angew. Chem. Int. Ed.* **50**, 9907–9910 (2011).
52. P. Mao, D. Wang, Top-down proteomics of a drop of blood for diabetes monitoring. *J. Proteome Res.* **13**, 1560–1569 (2014).
53. Y. Chen, P. Mao, D. Wang, Quantitation of intact proteins in human plasma using top-down parallel reaction monitoring-MS. *Anal. Chem.* **90**, 10650–10653 (2018).
54. D. Bloemberg, J. Quadrilatero, Rapid determination of myosin heavy chain expression in rat, mouse, and human skeletal muscle using multicolor immunofluorescence analysis. *PLoS One* **7**, e35273 (2012).
55. R. Bottinelli, S. Schiaffino, C. Reggiani, Force-velocity relations and myosin heavy chain isoform compositions of skinned fibres from rat skeletal muscle. *J. Physiol.* **437**, 655–672 (1991).
56. Z. R. Gregorich *et al.*, Top-down targeted proteomics reveals decrease in myosin regulatory light-chain phosphorylation that contributes to sarcopenic muscle dysfunction. *J. Proteome Res.* **15**, 2706–2716 (2016).
57. Z. Lin *et al.*, Simultaneous quantification of protein expression and modifications by top-down targeted proteomics: A case of the sarcomeric subproteome. *Mol. Cell. Proteomics* **18**, 594–605 (2019).
58. Z. R. Gregorich, Y. Ge, Top-down proteomics in health and disease: Challenges and opportunities. *Proteomics* **14**, 1195–1210 (2014).
59. Y. Jin *et al.*, Comprehensive analysis of tropomyosin isoforms in skeletal muscles by top-down proteomics. *J. Muscle Res. Cell Motil.* **37**, 41–52 (2016).
60. B. Wei, J.-P. Jin, Troponin T isoforms and posttranscriptional modifications: Evolution, regulation and function. *Arch. Biochem. Biophys.* **505**, 144–154 (2011).
61. T. Tucholski *et al.*, Distinct hypertrophic cardiomyopathy genotypes result in convergent sarcomeric proteoform profiles revealed by top-down proteomics. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 24691–24700 (2020).
62. S. K. Sze, Y. Ge, H. Oh, F. W. McLafferty, Top-down mass spectrometry of a 29-kDa protein for characterization of any posttranslational modification to within one residue. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 1774–1779 (2002).
63. J. B. Shaw *et al.*, Complete protein characterization using top-down mass spectrometry and ultraviolet photodissociation. *J. Am. Chem. Soc.* **135**, 12646–12651 (2013).

64. C. Lombard-Banek, S. A. Moody, P. Nemes, Single-cell mass spectrometry for discovery proteomics: Quantifying translational cell heterogeneity in the 16-cell frog (*Xenopus*) embryo. *Angew. Chemie Int. Ed.* **55**, 2454–2458 (2016).
65. R. L. Moss, G. M. Diffee, M. L. Greaser, "Contractile properties of skeletal muscle fibers in relation to myofibrillar protein isoforms BT" in *Reviews of Physiology, Biochemistry and Pharmacology, Volume 126*" (Springer, Berlin Heidelberg, 1995), pp. 1–63.
66. F. Lang *et al.*, Single muscle fiber proteomics reveals distinct protein changes in slow and fast fibers during muscle atrophy. *J. Proteome Res.* **17**, 3333–3347 (2018).
67. Y. Jin *et al.*, Complete characterization of cardiac myosin heavy chain (223 kDa) enabled by size-exclusion chromatography and middle-down mass spectrometry. *Anal. Chem.* **89**, 4922–4930 (2017).
68. X. Han, M. Jin, K. Breuker, F. W. McLafferty, Extending top-down mass spectrometry to proteins with masses greater than 200 kilodaltons. *Science* **314**, 109–112 (2006).
69. A. S. Deshmukh *et al.*, Deep proteomics of mouse skeletal muscle enables quantitation of protein isoforms, metabolic pathways, and transcription factors. *Mol. Cell. Proteomics* **14**, 841–853 (2015).
70. A. S. Deshmukh *et al.*, Deep muscle-proteomic analysis of freeze-dried human muscle biopsies reveals fiber type-specific adaptations to exercise training. *Nat. Commun.* **12**, 1–15 (2021).
71. W. Guo, M. L. Greaser, "Chapter 2–Muscle structure, proteins, and meat quality" in *New Aspects of Meat Quality*, P. P. Purslow, Ed. (Woodhead Publishing Series in Food Science, Technology and Nutrition, Woodhead Publishing, 2022), pp. 15–37.
72. M. Zhou *et al.*, Sensitive top-down proteomics analysis of a low number of mammalian cells using a nanodroplet sample processing platform. *Anal. Chem.* **92**, 7087–7095 (2020).
73. K. R. Johnson, Y. Gao, M. Greguš, A. R. Ivanov, On-capillary cell lysis enables top-down proteomic analysis of single mammalian cells by CE-MS/MS. *Anal. Chem.* **94**, 14358–14367 (2022).
74. W. Cai *et al.*, Top-down proteomics of large proteins up to 223 kDa enabled by serial size exclusion chromatography strategy. *Anal. Chem.* **89**, 5467–5475 (2017).
75. T. N. Tiambeng, Z. Wu, J. A. Melby, Y. Ge, "Size exclusion chromatography strategies and MASH explorer for large proteoform characterization" in *Proteoform Identification: Methods and Protocols*, L. Sun, X. Liu, Eds. (Springer, US, 2022), pp. 15–30.
76. S. G. Valeja *et al.*, Three dimensional liquid chromatography coupling ion exchange chromatography/hydrophobic interaction chromatography/reverse phase chromatography for effective protein separation in top-down proteomics. *Anal. Chem.* **87**, 5363–5371 (2015).
77. J. C. Tran *et al.*, Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature* **480**, 254–258 (2011).
78. G. M. Diffee, M. L. Greaser, F. C. Reinach, R. L. Moss, Effects of a non-divalent cation binding mutant of myosin regulatory light chain on tension generation in skinned skeletal muscle fibers. *Biophys. J.* **68**, 1443–1452 (1995).
79. K. A. Edman, The velocity of unloaded shortening and its relation to sarcomere length and isometric force in vertebrate muscle fibres. *J. Physiol.* **291**, 143–159 (1979).
80. Z. Wu *et al.*, MASH explorer: A universal software environment for top-down proteomics. *J. Proteome Res.* **19**, 3867–3876 (2020).