TECHNICAL BRIEF

Label-free profiling of skeletal muscle using high-definition mass spectrometry

Jatin G. Burniston[1], Joanne Connolly[2], Heikki Kainulainen[3], Steven L. Britton[4] and Lauren G. Koch[4]

[1] Research Institute for Sport and Exercise Sciences, Liverpool John Moores University, Liverpool, UK
[2] Waters Inc., Manchester, UK
[3] Department of Biology of Physical Activity, University of Jyväskylä, Jyväskylä, Finland
[4] Department of Anesthesiology, University of Michigan, Ann Arbor, MI, USA

Received: April 1, 2014
Revised: June 10, 2014
Accepted: July 22, 2014

We report automated and time-efficient (2 h per sample) profiling of muscle using ultra-performance LC coupled directly with high-definition MS (HDMS^E). Soluble proteins extracted from rat gastrocnemius ($n = 10$) were digested with trypsin and analyzed in duplicate using a 90 min RPLC gradient. Protein identification and label-free quantitation were performed from HDMS^E spectra analyzed using Progenesis QI for Proteomics software. In total 1514 proteins were identified. Of these, 811 had at least three unique peptides and were subsequently used to assess the dynamic range and precision of LC-HDMS^E label-free profiling. Proteins analyzed by LC-HDMS^E encompass the entire complement of glycolytic, β-oxidation, and tricarboxylic acid enzymes. In addition, numerous components of the electron transport chain and protein kinases involved in skeletal muscle regulation were detected. The dynamic range of protein abundances spanned four orders of magnitude. The correlation between technical replicates of the ten biological samples was $R^2 = 0.9961 \pm 0.0036$ (95% CI = 0.9940 – 0.9992) and the technical CV averaged 7.3 ± 6.7% (95% CI = 6.87 – 7.79%). This represents the most sophisticated label-free profiling of skeletal muscle to date.

Keywords:
Animal proteomics / Data-independent acquisition / Ion mobility / LC-MS

Additional supporting information may be found in the online version of this article at the publisher's web-site

Skeletal muscle accounts for ~40% of adult body mass and is the major site of both protein turnover and insulin-mediated glucose disposal. As such, abnormalities in muscle insulin sensitivity and protein metabolism contribute to the pathogenesis of diseases, such as type 2 diabetes and age-related declines in physical function. Conversely, the protective effects of exercise against noncommunicable diseases and frailty are in part underpinned by muscle adaptations. Therefore, proteome profiling of physiological and pathophysiological muscle adaptations may discover clinically relevant biomarkers. Despite its importance, skeletal muscle is a relatively understudied area in proteomics, somewhat due to the relative difficulty in achieving deep analysis of the muscle proteome. Thus, there remains a need for a reproducible and time-efficient method for screening muscle samples.

Currently, the majority of muscle proteome data comes from 2DE (e.g. [1]), which affords robust comparative analysis of protein species but does not easily resolve proteins at the extremes of the M_r and pI ranges. Moreover, because 2DE separates proteins to their constituent species, the number of nonredundant proteins identified may be relatively small (e.g. <300 proteins). In contrast, orthogonal separation by SDS-PAGE and RPLC coupled to ESI MS/MS (i.e. GeLC-MS/MS) is able to catalogue larger numbers of muscle proteins. For

Correspondence: Dr. Jatin G. Burniston, Muscle Physiology and Proteomics Laboratory, Research Institute for Sport and Exercise Sciences, Liverpool John Moores University, Tom Reilly Building, Byrom Street, Liverpool, L3 3AF, UK
E-mail: j.burniston@ljmu.ac.uk
Fax: +44-0-151-904 6284

Abbreviations: DIA, data-independent acquisition; **HDMS^E**, high-definition MS; **QIP**, QI for Proteomics; **RA**, relative abundance; **SILAM**, stable isotope labeled amino acids in mammals

instance, Højlund et al. [2] identified 945 proteins encompassing the entire complement of metabolic enzymes and some regulatory kinases. However, the number of proteins that can be differentially profiled using label-free GeLC-MS/MS is much fewer. Primarily, this is because the extensive sample fractionation makes it difficult to measure equivalent proteins or peptides in each biological replicate. At present, the largest number of skeletal muscle proteins to be differentially profiled by GeLC-MS/MS is 438, reported in [3], wherein spectral counting was used to investigate responses to aerobic exercise in patients with type 2 diabetes. Interestingly, Parker et al. [4] reports that the efficiency of data-dependent acquisition MS/MS to identify muscle proteins is essentially consistent across three different experimental designs (i.e. 1D LC-MS/MS, 2D LC-MS/MS, and GeLC-MS/MS). This suggests that the mass spectrometer may be a limiting factor in the analysis of the muscle proteome.

Herein, we used a Q-TOF mass spectrometer incorporating an additional gas-phase separation (i.e. ion mobility separation; IMS), which improves the resolution of complex peptide mixtures [5]. In addition, data-independent acquisition (DIA) was used whereby the instrument was programmed to alternate between low-energy (MS) and elevated-energy (MS^E) modes to record parent- and product-ion data, respectively. Finally, we performed the first label-free profiling of muscle using the recently released Progenesis QI for Proteomics (QIP) application (Nonlinear Dynamics, Newcastle, UK) that has been specifically developed to combine the strengths of IMS and DIA, that is, $HDMS^E$. QIP accurately aligns data to a common reference chromatogram so there are no missing values, and samples are normalized using a global scaling factor that improves the S/N over ratiometric scaling to a spiked purified protein (e.g. "hi-3" method advocated in [6]). We combined these advances in MS and bioinformatics with 1D nanoflow RP UPLC to create a time-efficient and robust technique for routine comparative analysis of skeletal muscle samples.

Gastrocnemius muscles were obtained from genetically heterogeneous N:NIH rats (described in [7]) and in-solution trypsin digestion was performed according to [8]. Peptide mixtures were analyzed by ultra performance LC (nano ACQUITY, Waters, Milford, MA) and online IMS MS (SYNAPT G2-S, Waters, Manchester, UK). Samples (200 ng tryptic peptides) were loaded in aqueous 0.1% v/v formic acid via a Symmetry C_{18} 5 μm, 2 cm × 180 μm trap column (Waters). Separation was conducted at 35°C through an HSS T3 C_{18} 1.8 μm, 25 cm × 75 μm analytical column (Waters). Peptides were eluted using a gradient rising to 40% acetonitrile 0.1% v/v formic acid over 90 min at a flow rate of 300 nL/min. Additionally, a Lockmass reference (100 fmol/ μL Glu-1-fibrinopeptide B) was delivered to the NanoLockSpray source of the mass spectrometer at a flow rate of 500 nL/ min, and was sampled at 60 s intervals.

For all measurements, the mass spectrometer was operated in a positive ESI mode at a resolution of >25 000 full-width at half maximum (FWHM). Prior to analysis, the TOF analyzer was calibrated with a NaCsI mixture from m/z 50 to 1990. $HDMS^E$ analyses were conducted within the Triwave ion guide. Accumulated ions were separated according to their drift time characteristics in the N_2 gas-filled mobility cell prior to CID alternating between low (4 eV) and elevated (14–40 eV) collision energies at a scan speed of 0.9 s per function over 50–2000 m/z.

Analytical data were LockMass corrected postacquisition using the doubly charged monoisotopic ion of the Glu-1-fibrinopeptide B. Charge reduction and deconvolution of potential parent-fragment correlation was achieved in the first instance by means of retention and drift time alignment, as described previously [9].

$HDMS^E$ spectra were aligned using Progenesis QIP. Prominent ion features (approximately 1200 per chromatogram) were used as vectors to match each dataset to a common reference chromatogram. An analysis window of 10 min – 100 min and 50 m/z – 1650 m/z was selected, which encompassed a total of 47 109 features (charge states of +2, +3 or +4) and 3924 of these features were separated by IMS. Protein identifications and quantitative information were extracted using the dedicated algorithms in ProteinLynx GlobalSERVER (PLGS) v3.0 (Waters). Peak lists were searched against the Uniprot database (date 2 July, 2012) restricted to "Rattus" (7500 entries). The initial ion-matching requirements were ≧1 fragment per peptide, ≧3 fragments per protein and ≧1 peptide per protein. The enzyme specificity was trypsin allowing one missed cleavage, carbamidomethyl of cysteine (fixed) and oxidation of methionine (variable). Parent- and fragment-ion parts per million errors were calculated empirically and decoy databases were used to calculate the identification error rate. Scoring of the database searches was refined by correlation of physicochemical properties of fragmented peptides from theoretical and experimental data. Peptide identifications were imported to Progenesis QIP and filtered to exclude peptides with scores less than 5.5 [10]. In total, 16 749 peptides were identified and 1,018 had been resolved by IMS.

Proteins were grouped if they were identified by similar sets of peptides and, in total, $HDMS^E$ identified 1514 protein groups. Of these, 811 had three or more unique peptides and were subsequently used to investigate the breadth and precision of label-free profiling. Metabolic enzymes were the most numerous gene ontology phrase and proteins in this subset encompassed the entire complement of glycolytic, β-oxidation, and tricarboxylic acid enzymes. In addition, numerous components of the electron transport chain and protein kinases important in skeletal muscle regulation were detected (Fig. 1A and B, see also the Supporting Information). Consistent with the fast-twitch phenotype of the gastrocnemius, the relative abundance (RA) of enzymes involved in glycolysis and the tricarboxylic acid cycle were an order of magnitude greater than enzymes of fatty acid β-oxidation. Furthermore, the ratio of heart-type lactate dehydrogenase B (RA = 63 302) and the muscle isoform of lactate dehydrogenase A (RA = 557 602) was approximately 1:8. Kinases involved in skeletal muscle regulation were also detected; the

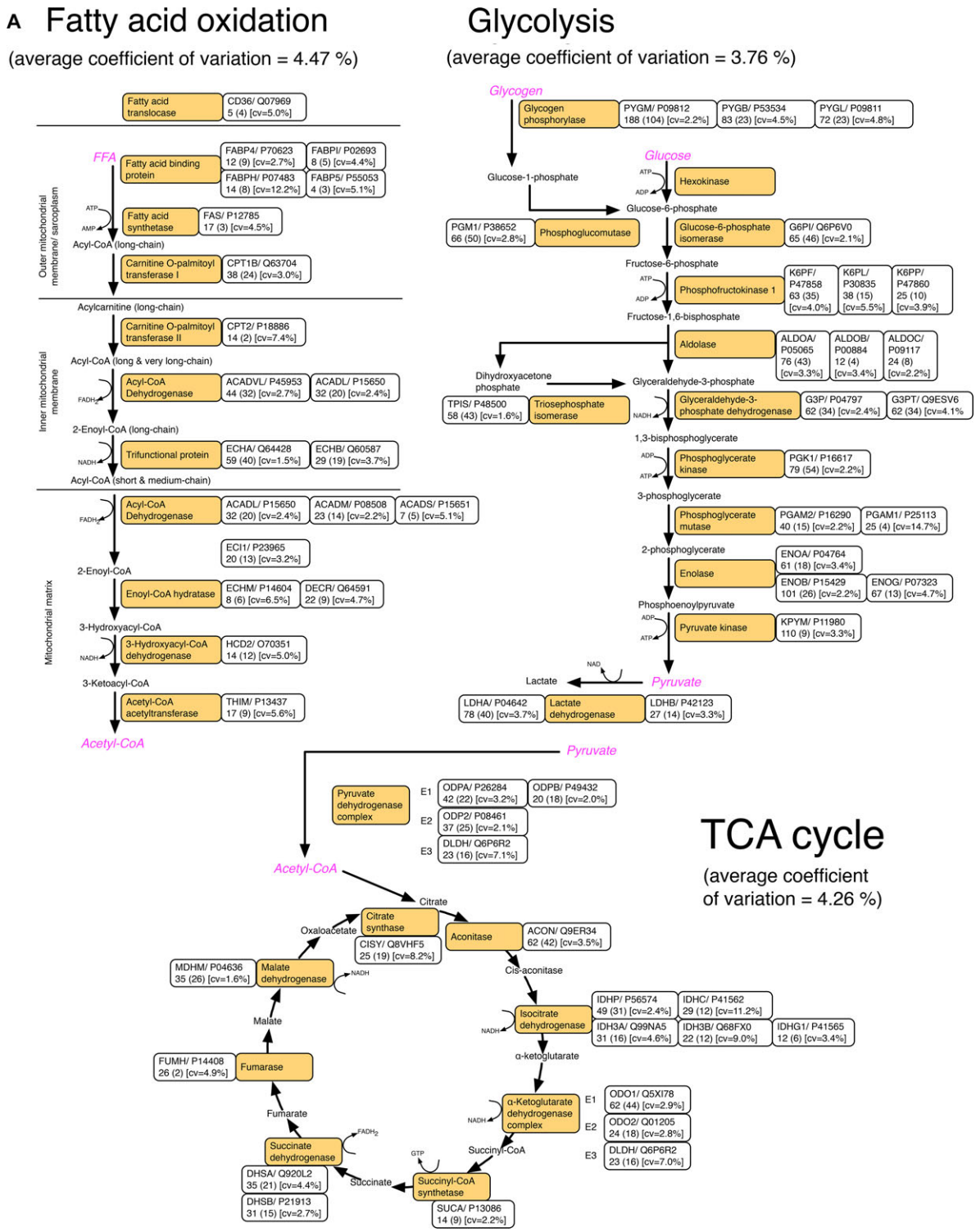

Figure 1. Profiling of soluble muscle proteins. The metabolic pathways of fatty acid β-oxidation, glycolysis, and the tricarboxylic acid cycle (A) are redrawn from the Kyoto Encyclopaedia of Genes and Genomes. For clarity, the respiratory chain (B) is not shown in its entirety, instead only subunits detected by HDMS^E profiling are highlighted. In addition, protein kinases (B) important in muscle regulation are displayed. Orange boxes display the common name of each enzyme, the adjacent boxes detail the UniProt protein ID and accession number, below these the total number of peptides is given. Numbers in parentheses represent the number of detected peptides that were unique to that protein and subsequently used for quantitative analyses. The technical CV for each protein is also given. (This figure is also available in the Supporting Information.)

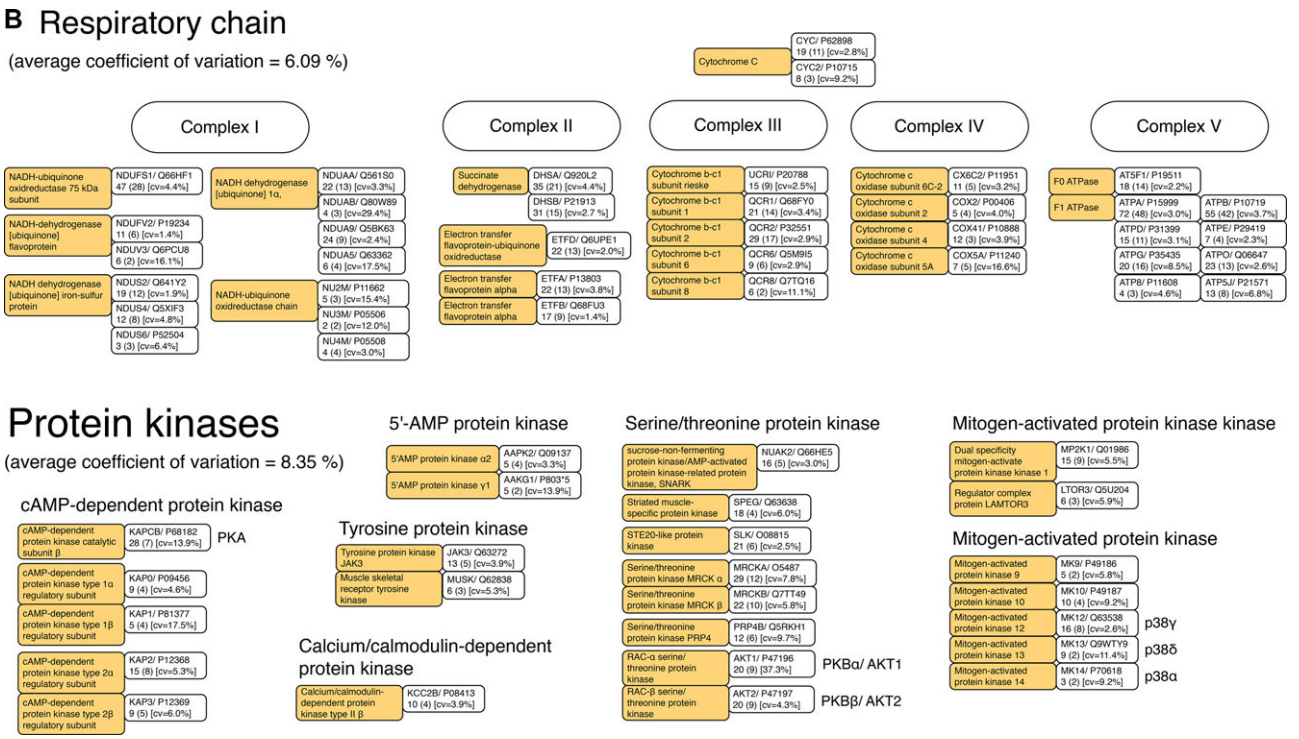

Figure 1. Continued.

most abundant (RA = 254 111) was sucrose-nonfermenting protein kinase/adenine monophosphate-activated protein kinase-related protein kinase, SNARK (Q66HE5). This protein is implicated in contraction-induced glucose transport in skeletal muscle [11], but is not currently listed as expressed in skeletal muscle in the Uniprot knowledge base. Consistent with the known order of abundance of p38 mitogen-activated protein kinase isoforms, p38γ (Q63538; RA = 25,332) was more abundant than p38δ (Q9WTY9; RA = 7,118) and p38α (P70618; RA = 2,171).

The dynamic range of protein abundance spanned four orders of magnitude (Fig. 2). There was a trend for variation to be greater in low abundance proteins (Fig. 3A), nonetheless, the correlation between technical replicates of ten biological samples was $R^2 = 0.9961 \pm 0.0036$ (95% CI = 0.9940 – 0.9992). The CV averaged 7.3 ± 6.7% (95% CI = 6.87 – 7.79%) and ~95% (767 of 811) of proteins had a technical CV of less than 20% (Fig. 3B). Herein, we used genetically heterogeneous animals, and when data were organized to encompass both biological and technical variability the average CV was 21.3 ± 15.9% (range 4.5–104.4%; Fig. 3C). The variability of 95% of proteins (770 proteins) was <54%, therefore, we estimate [12] that a sample size of $n = 9$ would be sufficient to detect a 40% difference ($\alpha = 0.05$, 80% power) in protein abundance between two independent groups.

The major strengths of our HDMS^E technique include its precision, time efficiency, and applicability to archived muscle including unlabeled human samples. LC-MS/MS analysis

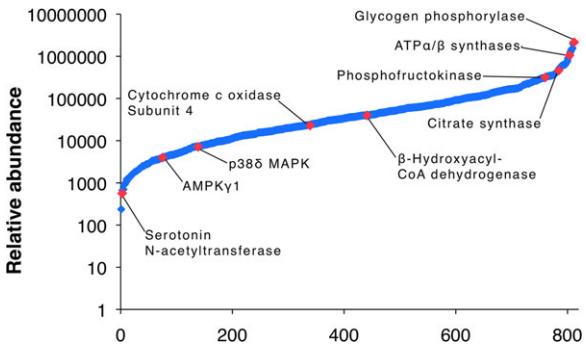

Figure 2. Dynamic range of HDMS^E profiling. The RA of proteins identified by HDMSE spanned four orders of magnitude. The most abundant protein was glycogen phosphorylase (2.1 × 10^6 AU). Consistent with the expected fast/glycolytic phenotype of gastrocnemius, glycolytic, and tricarboxylic acid enzymes were more abundant (~3.5e05 AU) than enzymes of fatty acid oxidation such as β-hydroxyacyl-CoA dehydrogenase (3.6e04 AU). Regulatory proteins, including mitogen-activated protein kinase p38δ and the gamma 1 subunit of AMP kinase were detected at an abundance of ~5.0 × 10^3.

of SILAC or stable isotope labeled amino acids in mammals (SILAM) is considered the gold-standard approach for differential proteome profiling. Differences between archetypal fast- and slow-twitch muscles have been investigated using the "SILAC" mouse [13], and recently 2474 proteins were identified [14] in mixed fiber muscles of these animals. This unprecedented level of muscle proteome coverage was

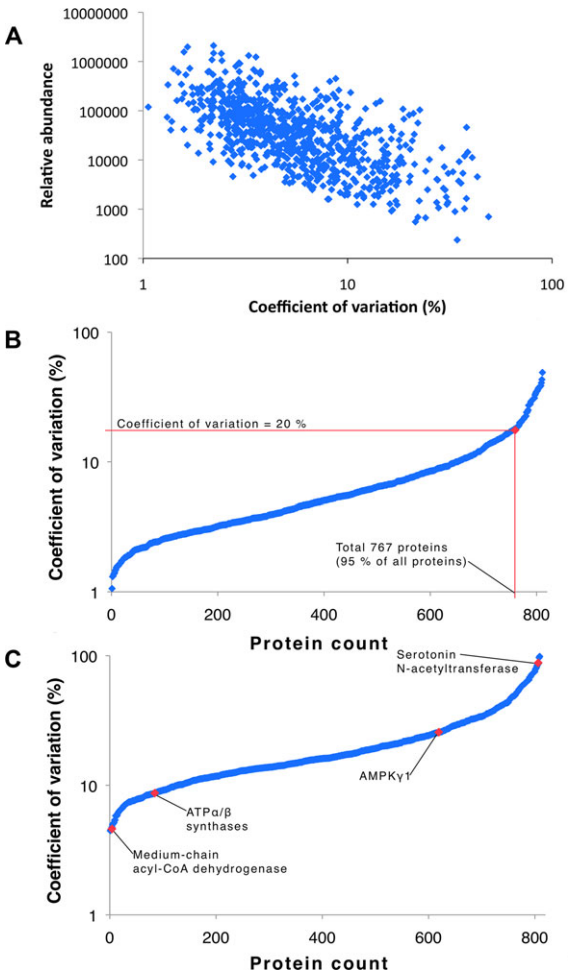

Figure 3. Technical and biological variability. The RA of a protein was inversely related to the level of technical variation (A). Technical CV (B) was normally distributed, and 95% (767 of 881) proteins identified had CV of less than 20%. Biological variability (range 4–104%; C) was greater than technical variation.

achieved by high-resolution LC-MS/MS of muscle digests separated into 12 fractionations by off-gel IEF. The correlation between replicate muscle samples was not provided but values of $R^2 = 0.93$ and 0.95 were reported from equivalent analysis of lung and liver proteins.

Here, we report a correlation of $R^2 > 0.99$ between technical replicates and an average CV of 7.3%. This high level of reproducibility may be due to our use of a simpler 1D RPLC and IMS separation strategy alongside DIA, which does not rely on stochastic selection of precursor ions so suffers less from inconsistent sampling of peptides. Moreover, because the instrument alternates between MS and MS^E this greatly improves the MS duty cycle compared to data-dependent acquisition and affords more detailed detection of peak profiles, further increasing measurement precision [10]. Each of these factors contributes toward enhancing reproducibility of Progenesis QIP analysis. The current profiling compares favorably against DIGE of skeletal muscle [15] where 70% of spots had a CV less than 14%. Using $HDMS^E$, the average CV of enzymes in the glycolytic, tricarboxylic acid cycle, and fatty acid β-oxidation pathways was 3.8, 4.3, and 4.5%, respectively (Fig. 1A and B, see also the Supporting Information). This level of precision is equivalent to spectrophotometric assays optimized for individual "key/rate-limiting" metabolic enzymes. Importantly, because $HDMS^E$ profiling simultaneously monitors all enzymes within each pathway, bioinformatic analysis of this highly parallel data may be used to discover novel interactions that correlate with changes in muscle (patho-) physiology.

Skeletal muscle is an accessible and clinically relevant tissue but proteomic analysis of muscle is challenging because approximately half of the total protein mass is occupied by just ten highly abundant contractile and metabolic proteins. Indeed, the dynamic range of muscle protein abundance is thought to be equivalent to that of the plasma proteome (i.e. spanning ~12 orders of magnitude). Our $HDMS^E$ profiling of soluble muscle proteins spanned four orders of magnitude in protein abundance, which is equivalent to DIGE analysis of similar samples in our hands [15] and GeLC-MS/MS analysis using spectral counting [3]. Orthogonal separation of SILAC peptides [14] achieved six orders of magnitude but required 20 h instrument time per sample. Similarly, label-free GeLC-MS/MS [3] required approximately 30 h per sample and profiled 438 proteins. Here, 1D RPLC of less than 2 h duration was used to analyze 811 soluble muscle proteins. Furthermore, this substantial increase in time efficiency and proteome coverage was achieved at a high level of precision.

In conclusion, we report a time-efficient method for precise profiling of >750 skeletal muscle proteins. Specifically, the current work represents a 20-fold increment in time efficiency and almost twofold improvement in proteome coverage compared to the equivalent exemplary label-free analyses using GeLC-MS/MS. Moreover, our technique does not require stable isotope labeling so can be applied to studies in both human and nonhuman animal models.

The authors have declared no conflict of interest.

References

[1] Burniston, J. G., Hoffman, E. P., Proteomic responses of skeletal and cardiac muscle to exercise. *Expert Rev. Proteomics* 2011, *8*, 361–377.

[2] Højlund, K., Yi, Z., Hwang, H., Bowen, B. et al., Characterization of the human skeletal muscle proteome by one-dimensional gel electrophoresis and HPLC-ESI-MS/MS. *Mol. Cell Proteomics* 2008, *7*, 257–267.

[3] Hussey, S. E., Sharoff, C. G., Garnham, A., Zhengping, Y. et al., Effect of exercise on the skeletal muscle proteome in patients with type 2 diabetes. *Med. Sci. Sports Exerc.* 2012, *45*, 1069–1076.

[4] Parker, K. C., Walsh, R. J., Salajegheh, M., Amato, A. A. et al., Characterization of human skeletal muscle biopsy samples

using shotgun proteomics. *J. Proteome Res.* 2009, *8*, 3265–3277.

[5] Shliaha, P. V., Bond, N. J., Gatto, L., Lilley, K. S., Effects of traveling wave ion mobility separation on data independent acquisition in proteomics studies. *J. Proteome Res.* 2013, *12*, 2323–2339.

[6] Silva, J. C., Gorenstein, M. V., Li, G. Z., Vissers, J. P., Geromanos, S. J., Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Mol. Cell Proteomics* 2006, *5*, 144–156.

[7] Koch, L. G., Pollott, G. E., Britton, S. L., A selectively bred rat model system for low and high response to exercise training. *Physiol. Genomics* 2013, *45*, 606–614.

[8] Malik, Z., Cobley, J., Morton, J., Close, G. et al., Label-free LC-MS profiling of skeletal muscle reveals heart-type fatty acid binding protein as a candidate biomarker of aerobic capacity. *Proteomes* 2013, *1*, 290–308.

[9] Li, G. Z., Vissers, J. P., Silva, J. C., Golick, D. et al., Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures. *Proteomics* 2009, *9*, 1696–1719.

[10] Levin, Y., Hradetzky, E., Bahn, S., Quantification of proteins using data-independent analysis (MSE) in simple and complex samples: a systematic evaluation. *Proteomics* 2011, *11*, 3273–3287.

[11] Koh, H. J., Toyoda, T., Fujii, N., Jung, M. M. et al., Sucrose nonfermenting AMPK-related kinase (SNARK) mediates contraction-stimulated glucose transport in mouse skeletal muscle. *Proc. Natl. Acad. Sci. USA* 2010, *107*, 15541–15546.

[12] van Belle, G., Millard, S. P., *STRUTS: Statistical Rules of Thumb*. Departments of Environmental Health and Biostatistics, University of Washington, Seattle, WA 1998.

[13] Drexler, H. C., Ruhs, A., Konzer, A., Mendler, L. et al., On marathons and Sprints: an integrated quantitative proteomics and transcriptomics analysis of differences between slow and fast muscle fibers. *Mol. Cell Proteomics* 2012, *11*, 1–16. M111.010801.

[14] Geiger, T., Velic, A., Macek, B., Lundberg, E. et al., Initial quantitative proteomic map of 28 mouse tissues using the SILAC mouse. *Mol. Cell Proteomics* 2013, *12*, 1709–1722.

[15] Burniston, J. G., Kenyani, J., Gray, D., Guadagnin, E. et al., Conditional independence mapping of DIGE data reveals PDIA3 protein species as key nodes associated with muscle aerobic capacity. *J. Proteomics* 2014, *106C*, 230–245.

