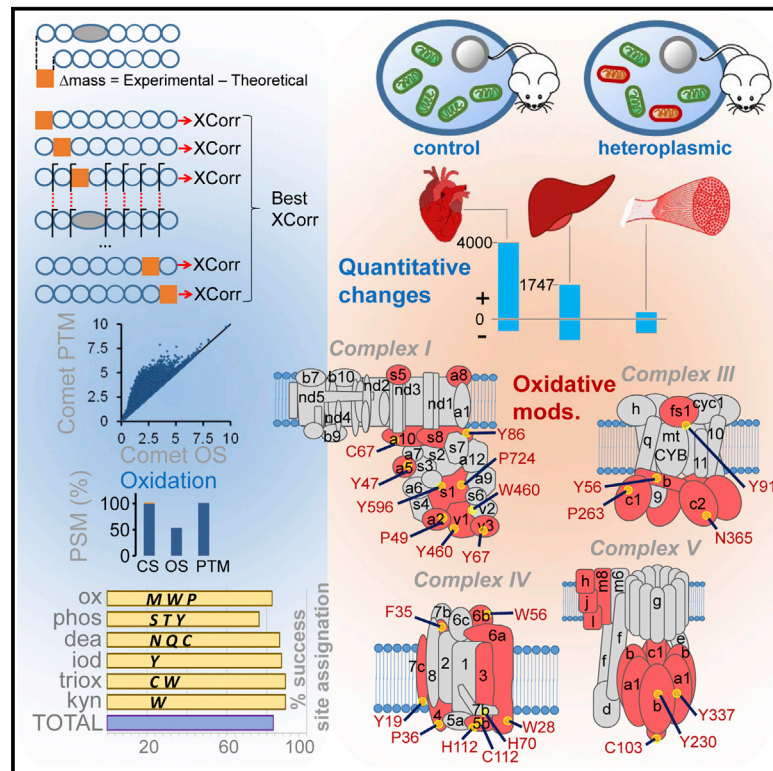


# Cell Reports

## Comprehensive Quantification of the Modified Proteome Reveals Oxidative Heart Damage in Mitochondrial Heteroplasmy

### Graphical Abstract



### Authors

Navratan Bagwan,  
Elena Bonzon-Kulichenko,  
Enrique Calvo, ..., Ana Latorre-Pellicer,  
José Antonio Enríquez, Jesús Vázquez

### Correspondence

jesus.vazquez@cnic.es

### In Brief

Bagwan et al. present a suite of algorithms for the unbiased identification and quantification of post-translational modifications and their site of modification by mass spectrometry. They illustrate its utility by showing that mitochondrial heteroplasmy in mice affects mainly the heart, inducing oxidative damage to proteins of the oxidative phosphorylation system.

### Highlights

- Algorithms for comprehensive identification of protein modifications by mass spectrometry
- Modified site is located with 85% accuracy
- Integrates quantitative analysis of the proteome and the modified peptidome
- mtDNA heteroplasmy causes oxidative damage in heart OXPHOS proteins



# Comprehensive Quantification of the Modified Proteome Reveals Oxidative Heart Damage in Mitochondrial Heteroplasmy

Navratan Bagwan,<sup>1,5</sup> Elena Bonzon-Kulichenko,<sup>1,2,5</sup> Enrique Calvo,<sup>1,2,5</sup> Ana Victoria Lechuga-Vieco,<sup>1,3</sup> Spiros Michalakopoulos,<sup>1</sup> Marco Trevisan-Herraz,<sup>1,2</sup> Iakes Ezkurdia,<sup>1,2</sup> José Manuel Rodríguez,<sup>1</sup> Ricardo Magni,<sup>1</sup> Ana Latorre-Pellicer,<sup>1</sup> José Antonio Enríquez,<sup>1,4</sup> and Jesús Vázquez<sup>1,2,6,\*</sup>

<sup>1</sup>Centro Nacional de Investigaciones Cardiovasculares Carlos III (CNIC), 28049 Madrid, Spain

<sup>2</sup>CIBER Cardiovascular Diseases (CIBERCV), Madrid, Spain

<sup>3</sup>CIBERES: C/ Melchor Fernández-Almagro 3, 28029 Madrid, Spain

<sup>4</sup>CIBERFES: C/ Melchor Fernández-Almagro 3, 28029 Madrid, Spain

<sup>5</sup>These authors contributed equally

<sup>6</sup>Lead Contact

\*Correspondence: [jesus.vazquez@cnic.es](mailto:jesus.vazquez@cnic.es)

<https://doi.org/10.1016/j.celrep.2018.05.080>

## SUMMARY

Post-translational modifications hugely increase the functional diversity of proteomes. Recent algorithms based on ultratolerant database searching are forging a path to unbiased analysis of peptide modifications by shotgun mass spectrometry. However, these approaches identify only one-half of the modified forms potentially detectable and do not map the modified residue. Moreover, tools for the quantitative analysis of peptide modifications are currently lacking. Here, we present a suite of algorithms that allows comprehensive identification of detectable modifications, pinpoints the modified residues, and enables their quantitative analysis through an integrated statistical model. These developments were used to characterize the impact of mitochondrial heteroplasmy on the proteome and on the modified peptidome in several tissues from 12-week-old mice. Our results reveal that heteroplasmy mainly affects cardiac tissue, inducing oxidative damage to proteins of the oxidative phosphorylation system, and provide a molecular mechanism explaining the structural and functional alterations produced in heart mitochondria.

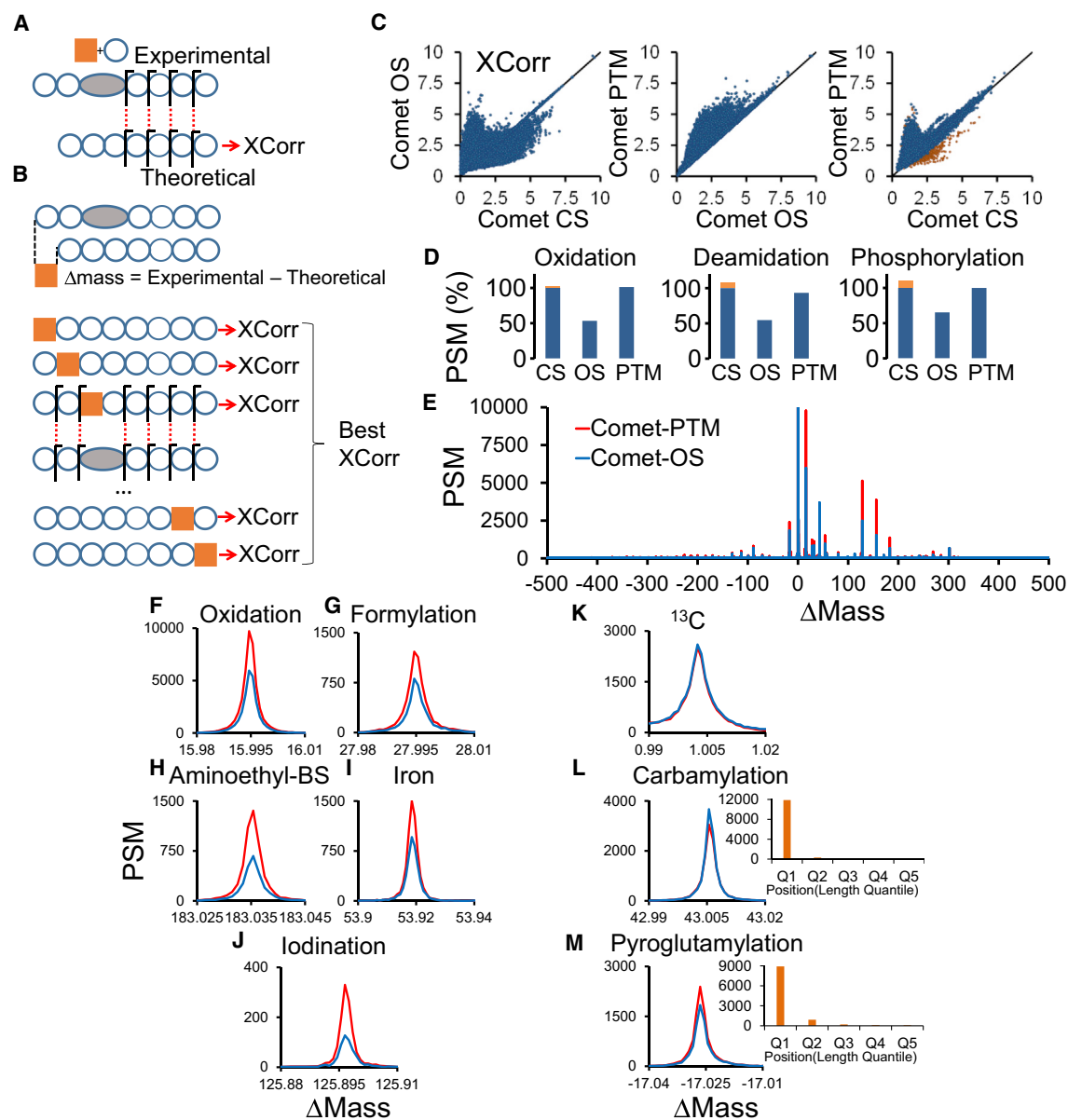
## INTRODUCTION

Shotgun mass spectrometry (MS)-based proteomics (Link et al., 1999) has become a powerful tool for biotechnological and biomedical research. Advances in speed and sensitivity allow the generation of millions of spectra per experiment, but only a minority of these spectra can be mapped to proteins (Griss et al., 2016; Skinner and Kelleher, 2015). A large proportion of unassigned spectra are thought to arise from peptides containing sequence variants or unknown chemical and post-translational modifications (PTMs) (Griss et al., 2016), and their charac-

terization is one of the most interesting and challenging goals in proteomics. A number of computational methods have been proposed for the detection of these unmatched peptides (Bern et al., 2007; Chen et al., 2009; Griss et al., 2016; Kim and Pevzner, 2014; Ma and Lam, 2014; Shortreed et al., 2015). Recently, an “open search” (OS) strategy, where precursor mass tolerances of hundreds of daltons were used with a conventional search engine, was reported to identify modified peptides at an unprecedented scale (Skinner and Kelleher, 2015). Another report demonstrated that OS can be performed at orders-of-magnitude faster speeds using a fragmentation-ion indexing algorithm (MSFragger) (Kong et al., 2017). These two methods may have a considerable impact on the field, opening the way to true hypothesis-free analysis of PTMs by MS; however, OS algorithms still rely on the chance that the modification leaves enough unaffected fragment ions for matching by the search engine (Figure 1A). OS strategies can therefore identify only approximately one-half of the modified peptides detectable by conventional “closed” searches (CS) (Chick et al., 2015). Moreover, existing OS approaches cannot directly identify the modification site. A further important consideration is that OS methods have not previously been used to quantify PTM alterations, and a general statistical model for the analysis of data of this kind is currently lacking.

Here, we present a suite of bioinformatics tools designed to overcome these limitations. Our tools double the coverage attained by existing algorithms, enabling the generation of comprehensive peptide maps that include practically all of the modifications potentially detectable by MS and CS of the database. Our approach also allows accurate location of the modified residues and quantitative analysis of PTMs in the context of proteome-wide studies. We demonstrate the performance of our tools by performing a comprehensive, tissue-specific characterization of PTMs induced by mitochondrial heteroplasmy in a mouse model. Heteroplasmy has recently attracted the attention of the biomedical community because it can be produced during mitochondrial replacement therapies aimed at preventing transmission of pathogenic mutations in mitochondrial DNA (Craven et al., 2010) or at treating infertility (Wolf





**Figure 1. Overview and Peptide Identification Performance of Comet-PTM**

(A) Conventional OS methods can identify modified peptides from MS/MS spectra, but only the fragments unaffected by the modification (orange square) are matched; this effect diminishes the score assigned to modified peptides, decreasing identification performance.

(B) Comet-PTM first calculates the difference between the mass of the candidate and the mass of the precursor ion detected by the MS ( $\Delta\text{mass}$ ).  $\Delta\text{mass}$  is then iteratively added to each amino acid in the peptide sequence, and the position that yields the best score is selected as the correct match.

(C) The HeLa dataset from the original OS article (Chick et al., 2015) was searched using Comet in closed search (CS) mode with three variable modifications (Met oxidation, Asn and Gln deamidation, and Ser and Thr phosphorylation), with Comet in OS mode, or with Comet-PTM (500-Da tolerance in the latter two cases). The scores obtained for the same spectra in the different searching conditions are compared. Yellow points in the rightmost graph are PSMs that match peptides with more than one modification in CS mode.

(D) Identification performance of OS and Comet-PTM relative to CS in the populations of peptides modified by oxidation, deamidation, or phosphorylation. The number of PSMs was obtained after filtering by the score threshold corresponding to 1% FDR in the CS. PSMs matching peptides with more than one modification in CS are indicated by orange bars.

(E) Frequency distribution of PSMs obtained by OS and Comet-PTM as a function of  $\Delta\text{mass}$ .

(F–M) Details of the frequency distribution around oxidation (F), formylation (G), aminoethylbenzenesulfonylation (H), iron (I), iodination (J),  $^{13}\text{C}$  (K), carbamylation (L), and pyroglutamylation (M). The charts to the right of (L) and (M) show the distribution of PSMs as a function of the quantile position of the modification in the peptide sequence assigned by Comet-PTM.

See also Figure S1.

et al., 2015). Mitochondrial heteroplasmy in mice can be genetically unstable and produce adverse physiological effects (Sharpley et al., 2012). The exact molecular mechanisms that produce the pathological effects are unclear and the potential health risk produced by heteroplasmy in the offspring is debated. Our results show that heteroplasmy between nonpathological mitochondrial DNA variants induces an array of oxidative modifications in the heart that predominantly affect proteins of the oxidative phosphorylation system. The tools presented here thus improve our ability to interpret the totality of information present in MS/MS datasets and provide proteome-wide perspectives for systems biology analysis in high-throughput proteomics.

## RESULTS

### Comet-PTM Enables Comprehensive Identification of Peptide Modifications

We developed Comet-PTM, an improved OS engine that takes into account the mass shift produced by the modification in the fragmentation series, producing the same score as a CS search using the same mass increment as a variable modification (Figure 1B). To test the performance of Comet-PTM, the results were compared with those obtained with OS and with CS using three common variable modifications. As expected, OS assigned higher scores to a large population of peptide-spectrum matches (PSMs) containing modifications not included in the CS list of variable modifications (Figure 1C, left). However, CS assigned a higher score than OS to a large population of PSMs with modifications included in the CS list, because the modifications affected the matching of fragment ions, decreasing the OS-assigned score (Figure 1C, left). This effect reduced the identification performance of OS to around 50% of the PTMs identified by CS (Figure 1D), confirming previous results (Chick et al., 2015). In contrast, scores obtained with Comet-PTM matched or exceeded OS (Figure 1C, center). In addition, Comet-PTM identified the same  $\Delta$ mass peaks as OS (Figure 1E); however, most of these peaks contained about 2-fold more PSMs (Figures 1E–1J), reflecting the superior performance. As expected, this effect was not observed when the modification did not affect the fragmentation series. Thus, the  $^{13}\text{C}$  peak, produced by errors in assignment of the monoisotopic precursor mass, was observed with exactly the same number of PSMs (Figure 1K). The number of PSMs was also similar for modifications in N-terminal position of the peptide, which only affect b-series and have a negligible effect on identification by higher-energy collisional dissociation (HCD) fragmentation (Figures 1L and 1M).

Comet-PTM produced scores similar to or higher than those obtained with CS (Figure 1C, right), except for a small population of peptides for which CS found two or more modifications (Figure 1C, orange dots). This effect was not due to differences between CS and Comet-PTM scores (Figures S1A and S1B). Therefore, and unlike OS, Comet-PTM had a similar identification performance to that of CS for the preselected modifications (Figure 1D). In several instances, Comet-PTM correctly located an oxidation on Trp, Pro, or Tyr that CS wrongly assigned to oxidation on Met, the predefined variable modification residue

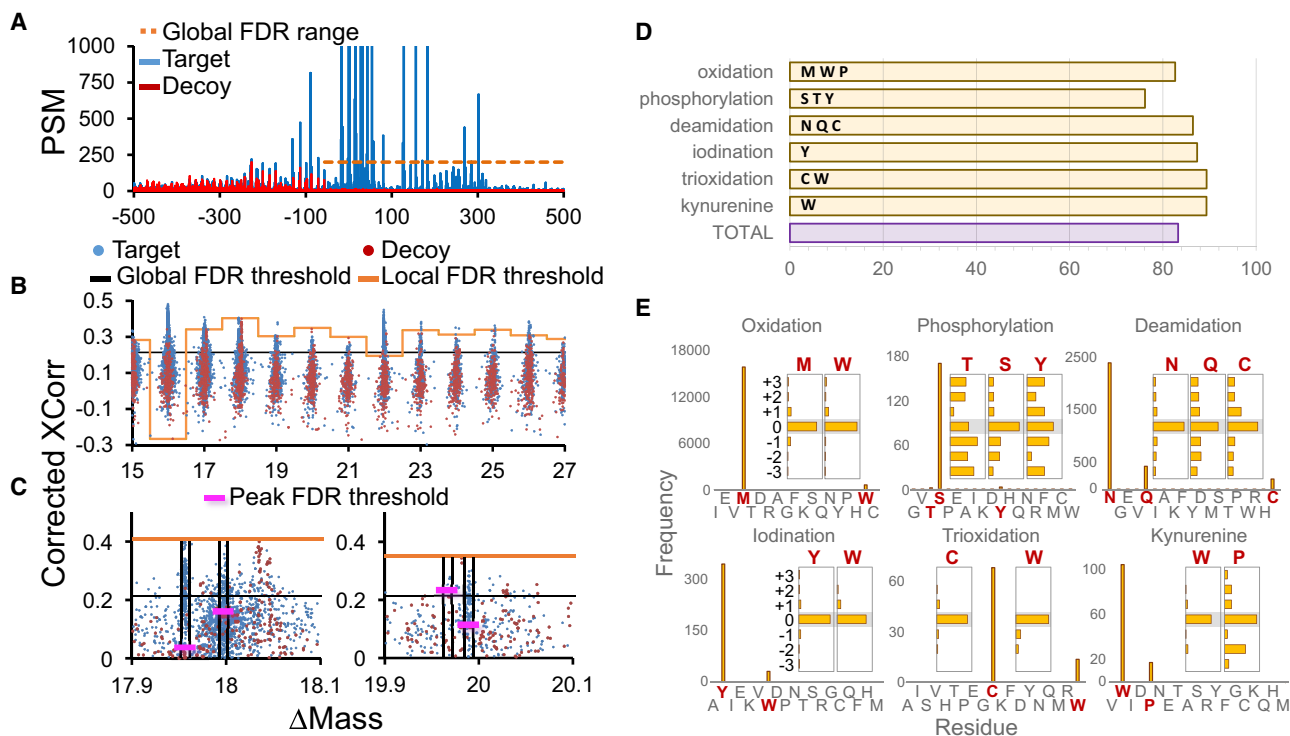
in CS (Figures S1C–S1E). This finding shows that conventional PTM searches using a variable modification on selected amino acids can force false assignment of the modified residue. Taken together, these data show that the Comet-PTM OS engine efficiently resolves the modification mass shift in the fragmentation series, doubling the number of modified peptides identified by conventional OS and matching the identification performance of targeted CS.

### Comet-PTM Detects the Location of Modifications in the Peptide Sequence

Peptides were identified after Comet-PTM searches with SHIFTS, an algorithm that detects the peaks in the  $\Delta$ mass distribution and controls the peptide false-discovery rate (FDR) through a conservative, three-layered approach (Figures 2A–2C; see also STAR Methods). From the output of a search against a concatenated target-decoy database, SHIFTS calculates a global score threshold to control FDR in the population of PSMs with  $\Delta$ mass greater than  $-56$  Da (Figure 2A). A local score threshold is also defined to control FDR separately within each of the  $\sim 1$ -Da bins in the  $\Delta$ mass distribution (Figure 2B). Finally, a peak score threshold is calculated to control FDR separately within each peak detected in the  $\Delta$ mass distribution (Figure 2C). A PSM is considered to be correctly identified when its score is above the global and peak thresholds; when a PSM does not form part of a peak, instead of the peak threshold the local threshold is used. This conservative approach allows full control of FDR, avoiding any bias due to the specific behavior of certain kinds of peptide modifications that may be more prone to match decoy sequences.

An advantage of Comet-PTM over existing OS approaches is that it automatically assigns the modification to the residue in the peptide sequence that produces the best score and therefore is the modified position that best explains the fragmentation data. Assigning modifications to specific residues is considered a much less reliable process than identifying peptides, partly because there is frequently insufficient information to determine the exact modified residue (Chalkley et al., 2008). To estimate the accuracy of Comet-PTM assignment of a modification to the correct site, we filtered correct identifications using the conservative method described above with a 1% FDR threshold and then calculated the fraction of PSMs in which the mass shift was located in amino acids predicted to harbor well-known modifications. Oxidation was located to Met, Trp, or Pro in 82% of cases, and deamidation to Asn, Gln, or carbamidomethylated Cys in 86% of cases (Figure 2D). Considering six well-known modifications, the overall accuracy of Comet-PTM was estimated at  $\sim 85\%$ . To test more specifically the performance of Comet-PTM for the correct assignment of phosphorylation sites, we analyzed a dataset obtained from synthetic phosphorylated peptides (Ferries et al., 2017). Comet-PTM assigned the correct position with 81% accuracy, and in 94% of the cases the modification was in the correct or in an adjacent position (Table S1).

We then analyzed which amino acids were assigned to known modifications with a higher frequency than expected by chance. Among the most frequent, we only found well-known modifications like oxidized Met and Trp, deamidated Asn, Gln and carbamidomethylated Cys, phosphorylated Ser, trioxidized Cys and



**Figure 2. Identification of Modified Peptides and Location of the Modified Site**

(A)  $\Delta$ Mass distribution of target and decoy PSMs and range of application of the global score threshold. The vertical scale is enlarged to show the distribution of decoy PSMs.

(B) Local score thresholds applied in each  $\sim 1$ -Da bin in the  $\Delta$ mass distribution.

(C) Peak score thresholds applied to the peaks detected by SHIFTS in the  $\Delta$ mass distribution.

(D) Percentage of successful assignment of the indicated modification to the indicated residues.

(E) Inset horizontal bar graphs plot the frequency distributions of PSMs assigned by Comet-PTM to the indicated amino acids and modifications. Vertical bar graphs: frequency of PSMs assigned to the indicated amino acids after subtracting the average frequency of the three previous and the three subsequent positions. Residues that accumulate counts in the histogram at position 0 are labeled in red.

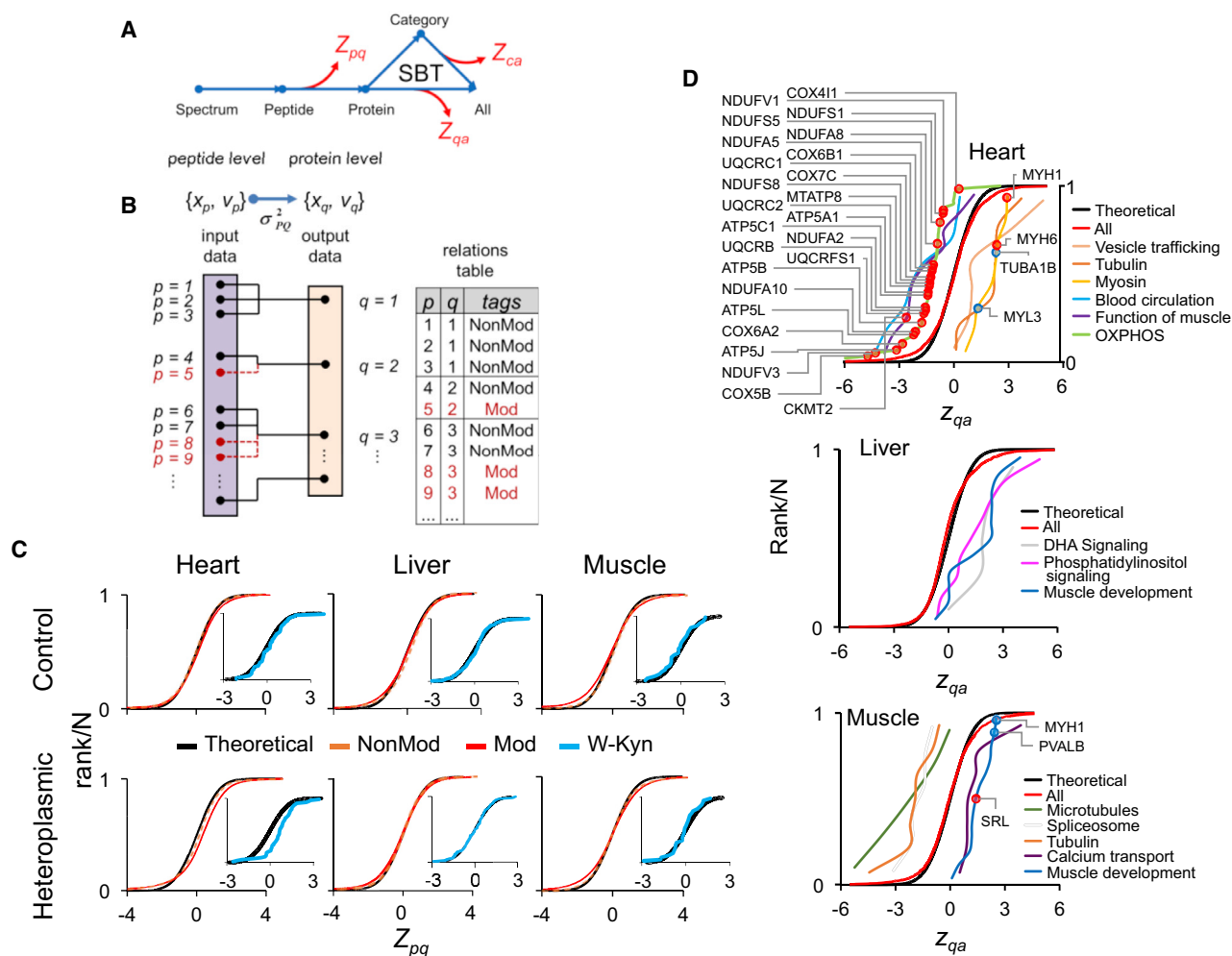
See also Figures S2 and S3.

Trp, and kynurenine-modified Trp (Figure 2E). Some Pro modifications were also found with the  $\Delta$ mass of kynurenine-modified Trp (Figure 2E), but all of them corresponded to homologous peptides that contained the Pro>Thr substitution, which has the same  $\Delta$ mass (Figure S2), and that were assigned the same score. Finally, Trp iodination was also clearly detected above background (Figure 2E). This modification was unexpected because, although Trp halogenases have been described in bacteria (van Pée and Patallo, 2006), they are not present in mammals, and Trp iodination has not been described before as a chemical artifact. The majority of peptides containing iodinated Trp contained neither Tyr nor His, the two amino acids known to react with iodine. A careful inspection of their fragment spectra confirmed the presence of  $\gamma$ -fragments that could only be explained by Trp being the modified residue (Figure S3). In addition, the only modification described in Unimod that matched the observed mass shift corresponded to iodination. From all of these data, we concluded that Trp was iodinated in these peptides, most probably as a side-chain reaction of the iodine produced from the Cys-alkylating reagent iodoacetamide. This finding further highlights the accuracy with which Comet-

PTM locates the site of modification and suggests that this OS engine may be useful for detecting novel peptide modifications in specific amino acids.

### A Single Integrated Statistical Framework Allows Quantification of the Proteome and of the Modified Peptidome

For the quantitative analysis of modified peptides, we developed an algorithm based on a previously proposed systematic workflow (García-Marqués et al., 2016; Navarro et al., 2014) (Figure 3A). The algorithm includes a peptide-to-protein integration step that quantifies protein values from the unmodified peptide forms and then computes the standardized log<sub>2</sub>-ratio of the modified peptides with respect to these protein values ( $Z_{pq}$ ) (Figure 3B). This makes it possible to detect modified peptides whose behaviors deviate significantly from those of the unmodified peptides from the same protein. Note that the algorithm calculates peptide abundances relative to the parent protein, not to the mean of all the peptides in the sample, so that changes at the peptide level are unaffected by changes in protein abundance. When applied to a biological model, the  $Z_{pq}$  distributions of the



**Figure 3. Quantitative Analysis of Alterations Produced by Heteroplasmy in the Modified Peptidome and in the Proteome**

(A) Scheme of the integrative workflow used to quantify modified peptides, proteins or categories. Each arrow represents a step performed with the generic integration algorithm (García-Marqués et al., 2016). Standardized  $\log_2$ -ratio values at the peptide level ( $Z_{pq}$ ) are obtained from the peptide-to-protein integration. The algorithm provides corrected peptide values by taking account of the corresponding protein changes. The same workflow is used for protein quantification and systems biology analysis of coordinated protein responses with the SBT model (García-Marqués et al., 2016).

(B) Scheme of the generic integration algorithm (GIA) (García-Marqués et al., 2016) adapted for the analysis of modified peptides in the peptide-to-protein integration step. The relations table include tags that can be used to indicate that modified peptides are excluded from the computation of protein averages.

(C) Distributions of  $Z_{pq}$  values for nonmodified (Non-Mod) and modified peptides (Mod). These data come from the analysis of mouse tissues in a model of mitochondrial heteroplasmy; each graph corresponds to a tissue from an individual animal. The inset plots show the  $Z_{pq}$  distributions for peptides containing the kynurenin Trp modification. Note that the distribution is displaced to the right in the hearts of heteroplasmic mice, indicating increased abundance. For further details, see STAR Methods.

(D) Characterization of coordinated protein alterations produced by heteroplasmy in heart, liver, and muscle using the SBT model.  $Z_{qa}$  values are standardized  $\log_2$ -ratio averages of proteins from heteroplasmic mice relative to controls. Proteins are grouped into representative functional categories with a statistically significant change. The upper panel presents proteins in the OXPHOS category that are decreased in the heteroplasmic heart and that contain heteroplasmy-induced oxidative modifications (Figure 6).

See also Table S2.

unmodified and modified peptide forms precisely followed the expected null hypothesis distribution in different tissues from two mouse strains (Figure 3C). These results convincingly demonstrate that the quantitative behavior of the entire population of modified peptides can be modeled accurately, providing a robust statistical framework within which to analyze abundance changes in high-throughput experiments. This allowed detection

of significant changes in certain modified peptides against the null hypothesis in specific situations (Figure 3C, lower left).

The same workflow used for peptide quantitation allowed the analysis of protein abundance changes and the characterization of functional categories that were affected by the coordinated action of proteins, using the systems biology triangle (SBT) model (García-Marqués et al., 2016) (Figure 3A). This statistical

design thus allowed a full description of the alterations in the modified peptidome in the global context of changes in the proteome.

### Heteroplasmy Produces Protein Alterations Consistent with a Mitochondrial Dysfunction in the Heart

We used Comet-PTM and associated tools to study the molecular impact of mitochondrial heteroplasmy on the proteome and the modified peptidome in heart, liver, and skeletal muscle from 12-week-old mice. This age was considered the most appropriate to study molecular alterations before more extensive damage was produced as a consequence of heteroplasmy. SBT model analysis of the quantitative protein data from the heart revealed a coordinated decrease of proteins related to muscle function and the oxidative phosphorylation system (OXPHOS) and a coordinated increase of tubulins, myosins, and proteins involved in vesicle trafficking. The muscle-function proteins included KCRS (mitochondrial creatine kinase) and VDAC3, which are key suppliers of phosphocreatine to the sarcomere, and SGCG and KGP1, which are essential for sarcomere contraction (Table S2). These alterations in heteroplasmic mice are highly consistent with the decreased ATP synthesis and the abnormal increase in phosphocreatine/ATP ratio in the heart and with the increase in plasma creatine kinase we have observed in this animal model (unpublished data), and provide evidence that heteroplasmic cardiac mitochondria have a compromised ability to supply energy to the sarcomere. Among the increased vesicle trafficking proteins, STXB3 and VAMP2 are implicated in the insulin-dependent movement of GLUT4 from inner vesicles to the plasma membrane. This finding also agrees with a higher glucose uptake we detected in heteroplasmic hearts, pointing to a shift toward glycolytic metabolism in order to compensate the reduced phosphocreatine supply from the OXPHOS. The coordinated decrease in blood proteins and the increase in heart cytoskeletal proteins in heteroplasmic animals suggest a homeostatic effort to maintain cardiac cell structure.

Heteroplasmy did not affect OXPHOS protein levels in liver but did produce a coordinated increase of docosahexaenoic acid (DHA) signaling pathway proteins (Figure 3D), including AKT2, BID, and B2CL1 (Table S2). AKT2 plays an essential role in the progression of the inflammatory response and apoptosis (López-Carballo et al., 2002); consistent with this finding, we detected signs of inflammation in the liver of heteroplasmic mice with aging (unpublished data). Heteroplasmic liver also showed coordinated increases in a group of myosins and phosphatidylinositol pathway proteins, probably reflecting a homeostatic response. These data thus indicate that, at 12 weeks of age, heteroplasmy produces a decrease in OXPHOS proteins in the heart, probably reflecting mitochondrial dysfunction, but the same effect is not seen in liver.

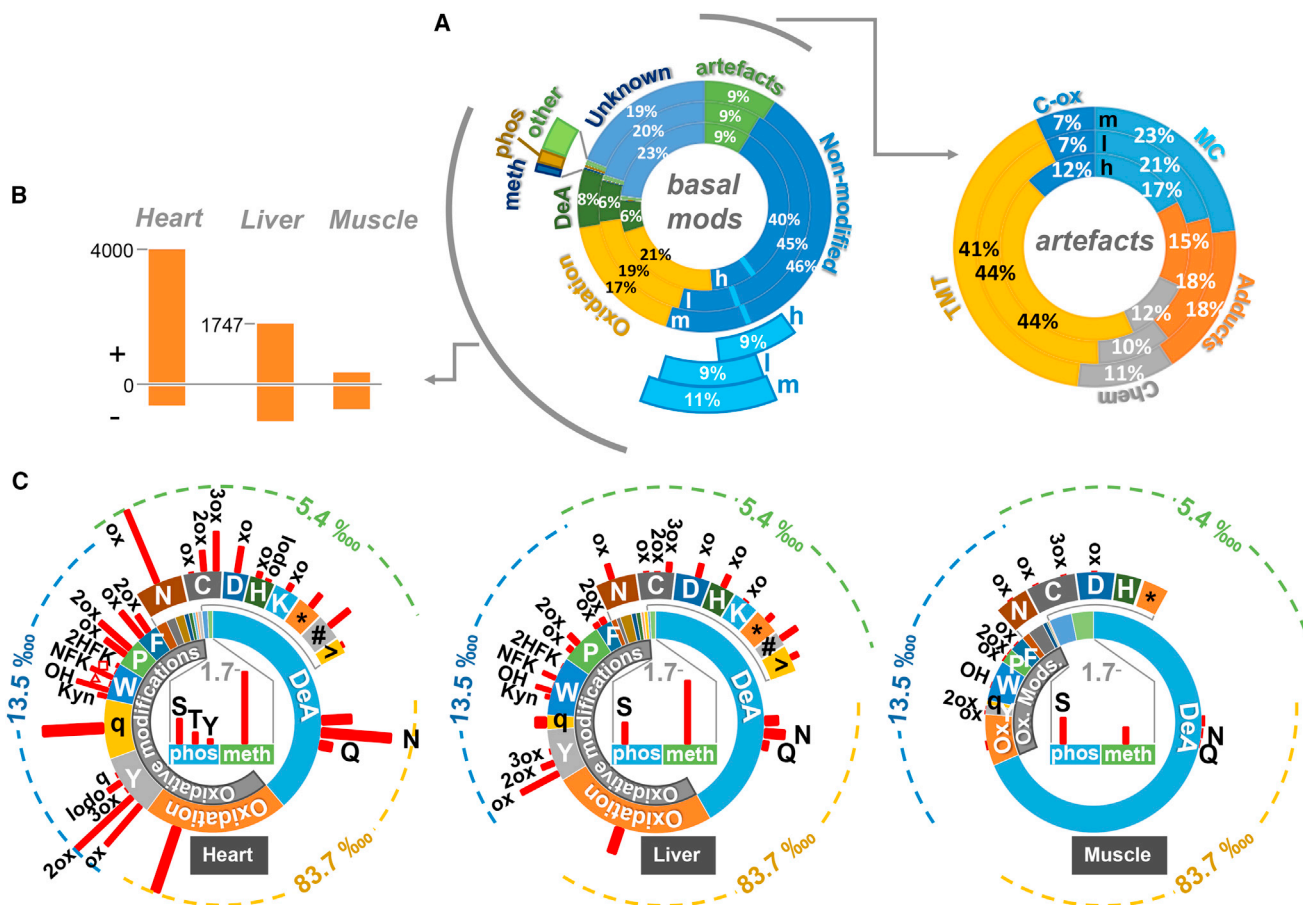
In heteroplasmic muscle, we detected a coordinated decrease of proteins regulating cell shape, such as tubulins and microtubule proteins (Figure 3D). On the other hand, the thick-filament myosins, as well as calcium transporting proteins, were increased, suggesting the induction of compensatory mechanisms. However, we found no evidence of altered mitochondrial protein levels in muscle at this age.

### Heteroplasmy Mainly Produces Oxidative Modifications of OXPHOS Proteins in Heart

To study the impact of heteroplasmy on the modified peptidome, we first compared the  $\Delta$ mass distribution of peptides searched with Comet-PTM and identified with SHIFTS in the three tissues. The distribution of the most abundant peaks was very similar in liver, heart, and muscle (Figure S4). In addition to the most intense peak corresponding to unmodified peptides, we found peaks corresponding to oxidized and deamidated peptides, and dioxidized peptides were also frequent. Other frequent modifications were artifacts produced by TMT reagents and iodoacetamide treatments, sodium adducts, and missed cleavages. Some tissue-specific modifications were also detected, mostly in the region from 50 to 100 Da. Among these, phosphorylated peptides were more abundant in muscle and 90% of these peptides belonged to proteins implicated in muscle contractility, consistent with the regulatory role of phosphorylation in the sarcomere. Within each  $\Delta$ mass peak, we analyzed the distribution of the modified amino acid residues, which were assigned to frequent modifications using the unbiased approach followed in Figure 2E. Confirming the results described above, amino acids known to harbor specific modifications were mostly located correctly (Figure S5). Interestingly, the relative proportion of modified sites was remarkably preserved across the three tissues, with the exception of tyrosine iodination, which was prominent in heart but less frequent in liver and skeletal muscle (Figure S5). This analysis also detected the increased frequency of phosphorylation in muscle (Figure S5). We estimated that about one-quarter of the total peptidome in the three tissues had known PTMs, mostly oxidation and deamidation; less than 10% of modifications were chemical artifacts (Figure 4A).

Second, we analyzed the quantitative data with the integrative statistical algorithm (Figure 3A) to determine the pattern of modifications that are specifically affected by heteroplasmy in each tissue. Heteroplasmy induced marked tissue-specific modifications, most of them concentrated in heart and barely detectable in skeletal muscle (Figure 4B; Table S3). The vast majority of increased modifications in heart were oxidative, affecting mostly Tyr, Trp, Pro, and Phe and to a lesser extent Asn, Cys, and Asp (Figure 4C). Heteroplasmy also induced phosphorylation and methylation in the heart. The modification pattern was similar in liver, but the effect of heteroplasmy in this tissue was significantly less pronounced.

Functional enrichment analysis revealed that heteroplasmy-induced modifications in mouse heart were mostly located in mitochondrial proteins, particularly those located in the inner mitochondrial membrane; moreover, most of the modified proteins were OXPHOS proteins, which were enriched with multiple oxidative modifications (Figure 5A). Tricarboxylic acid (TCA) cycle proteins and those related to the oxidoreductase complex were enriched in several modifications. Deamidation affected several mitochondria-related categories, and phosphorylation affected proteins related to contractility. Heteroplasmy-induced modifications affected considerably fewer categories in liver (Figure 5B) and were not significantly enriched in muscle (Figure 5C), where only a few categories were decreased, containing low numbers of proteins. Deamidated peptides were found in proteins related to mitochondria, NAD binding, and extracellular



**Figure 4. Alterations Produced by Heteroplasmy in the Modified Peptidome in Different Mouse Tissues**

(A) Complete map of the basal peptidome, expressed as a percentage of the total number of peptides identified in each tissue (circular bar graphs: h, heart; l, liver; m, muscle). For simplicity, peptides oxidized on Met (light blue) are included in the group of nonmodified peptides. The plot to the right shows proportions of artifacts produced by sample manipulation.

(B) Number of PTM-containing peptides significantly increased (+) or decreased (–) in heteroplasmic mouse tissues ( $n = 4$ ) relative to controls ( $n = 3$ ).

(C) Statistically significant PTM increases in heteroplasmic mouse tissues according to the type of modification and the modified residue. The circular inner bars show the peptide proportion for each modification in the basal state, and the radial red bars represent the proportion of peptides of each kind (in parts per 10,000) that are increased in heteroplasmic tissues. For the sake of clarity, three different scales are used depending on the type of modification. DeA, deamidation; meth, methylation; phos, phosphorylation; MC, missed cleavages; Chem, chemical derivatives produced by sample preparation; TMT, extra addition of the isobaric labeling reagent; Adducts, Na, K, and ammonia adducts. Uppercase letters indicate the single-letter amino acid code. ox, oxidation; 2ox, dioxidation; 3ox, trioxidation; Kyn, Trp to kynurenine; OH, Trp to hydroxytryptophan; NFK, Trp to N-formyl kynurenine; 2HFK, Trp to 2-hydroxy formyl kynurenine; red triangle, Trp to oxolactone; red square, Trp to quinone; q, quinone, Tyr to quinone; Iodo, iodination; asterisk, Met to homocysteic acid; hashtag, Met to homoserine.

See also [Figures S4](#) and [S5](#) and [Table S3](#).

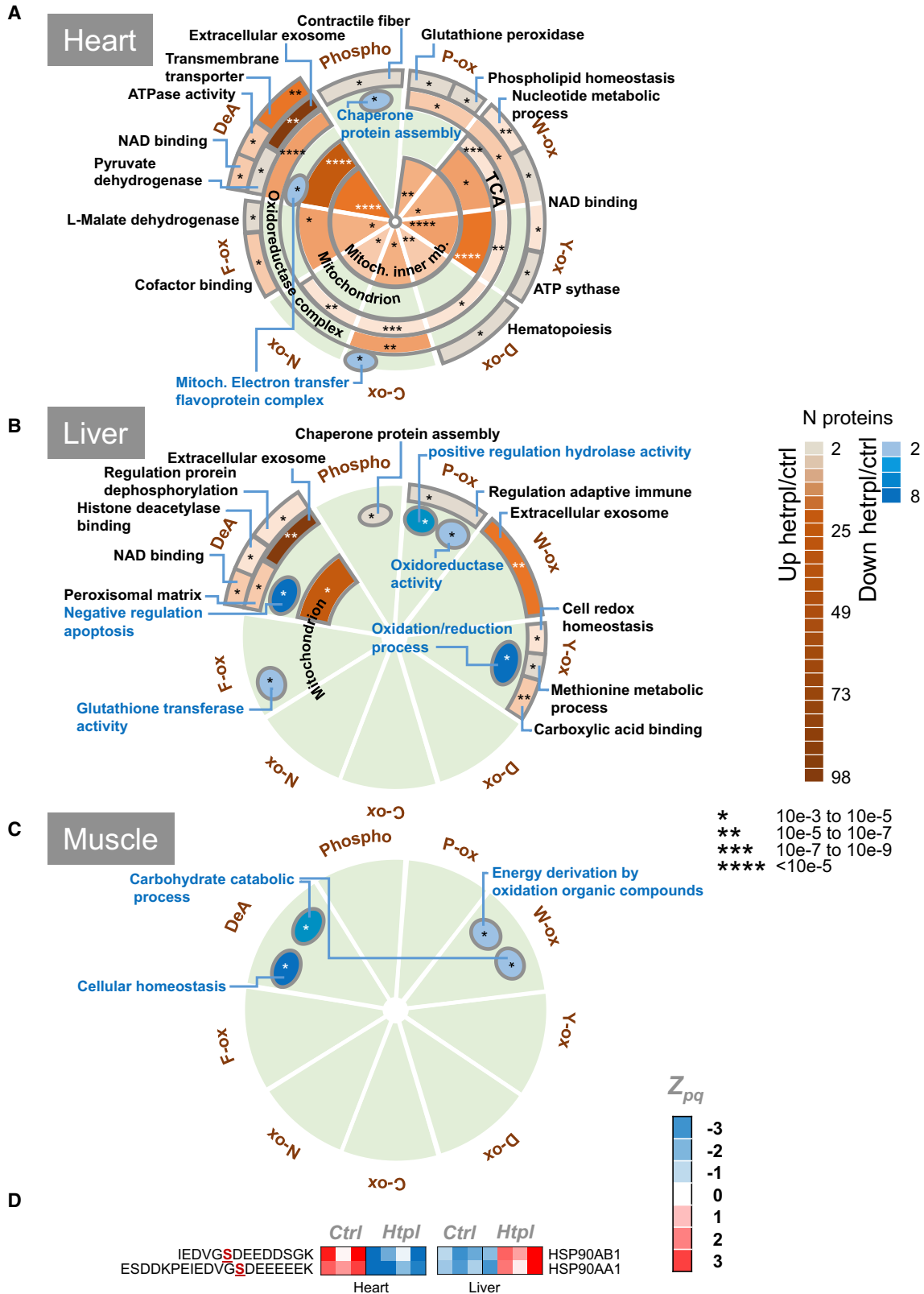
exosomes in both heart and liver, suggesting that these modifications do not contribute to tissue-specific phenotypic differences between heteroplasmic and control mice.

Interestingly, heteroplasmy decreased phosphorylations in proteins related to chaperone activity in the heart but had the opposite effect in liver ([Figures 5A](#) and [5B](#)). This category contains the two HSP90 isoforms (HSP90AA1 and HSP90AB1), which harbor PTM sites involved in the regulation of the chaperone activity ([Mollapour and Neckers, 2012](#)). These isoforms are known to be phosphorylated at Ser263 and Ser255, respectively ([Hu et al., 2015; Wang et al., 2017](#)), within the CK2 phosphorylation motif (S-X-X-E), which is crucial for protein activity. In particular, phosphorylation at Ser255 is required for the activa-

tion of mitogen-activated protein kinase (MAPK)/extracellular signal-regulated kinase (ERK) pathway. Comet-PTM and SHIFTS identified these two phosphorylated sites, and heteroplasmy decreased their levels in heart while increasing them in liver ([Figure 5D](#)). These findings suggest that heteroplasmy differentially affects chaperone activity and support the roles in cardiac heteroplasmy we have proposed for the endoplasmic reticulum and the mitochondrial unfolded stress response (unpublished data).

Heteroplasmy also altered phosphorylation levels at sites known to regulate protein activity in a tissue-specific manner. Phosphorylation at Ser42 from PTRF, which regulated transcriptional activity in response to metabolic challenges ([Huttlin et al.,](#)





(legend on next page)

2010; Liu and Pilch, 2016), was decreased in heart but was increased in muscle, in agreement with the different metabolic status of these tissues (Table S3). Similarly Ser133 from MYG, which regulates contractility (Lundby et al., 2013), was more phosphorylated in heart, in line with the coordinated decrease of proteins related to muscle function (Figure 3D), suggesting alterations in the contractile function of the heart in heteroplasmic animals. In the liver, the active site Ser117 from PGM1, which regulates glycogen biosynthesis (Huttlin et al., 2010; Lee et al., 2014), was less phosphorylated, suggesting lower fuel storage in this tissue (Table S3). These results reinforce the notion that the impact of heteroplasmy is tissue specific. The rest of phosphorylated sites affected by heteroplasmy had unknown regulatory roles, but it should be noted that all of them have been described previously, supporting the accuracy of the algorithms used to detect them. They include Ser181 in PGRMC1 (Jin et al., 2004), Ser442 in SRCA (Dai et al., 2007), Ser475 in JPH1 (Rigbolt et al., 2011), and Ser298 and Ser304 in BIN1 (Ballif et al., 2004; Huttlin et al., 2010) (Table S3).

To compare the alterations in the modified peptidome with the changes observed at the protein level, we mapped the proteins harboring PTMs affected by heteroplasmy onto the protein distribution plots depicting coordinated protein responses (red and blue dots in Figure 3D). Most increased modifications mapped onto the OXPHOS protein category in heart, which was decreased, whereas proteins belonging to other categories were mostly unaffected. This finding suggests that the oxidative damage produced by heteroplasmy in heart OXPHOS proteins induced their degradation, and is consistent with the increase in oxidative stress we have observed in heart slices, embryonic fibroblasts, and adult fibroblasts of heteroplasmic mice (unpublished data).

Most heteroplasmy-induced oxidative modifications in OXPHOS proteins occurred in the four respiratory protein complexes, except for Trp oxidations, which were mostly concentrated in C-IV (Figure 6). The modifications were clustered in groups of proteins located together in the structure of the complexes and predominantly affected proteins oriented toward the mitochondrial matrix, where proteins are more exposed to mitochondrial reactive oxygen species (ROS). In further analysis, we used ConSurf to calculate evolutionary conservation profiles for the affected residues (Ashkenazy et al., 2016) and mapped the modifications onto the known three-dimensional (3D) structures of the complexes, when available, or onto their predicted models (Meier and Söding, 2015). We found that 70% of the modified residues were conserved or interacted with conserved regions (Table S4). Moreover, most of the modifications were located on the protein surface, and those that were buried tended to have high conservation scores, suggesting they are important for structural stability or folding (Table S4). These findings sug-

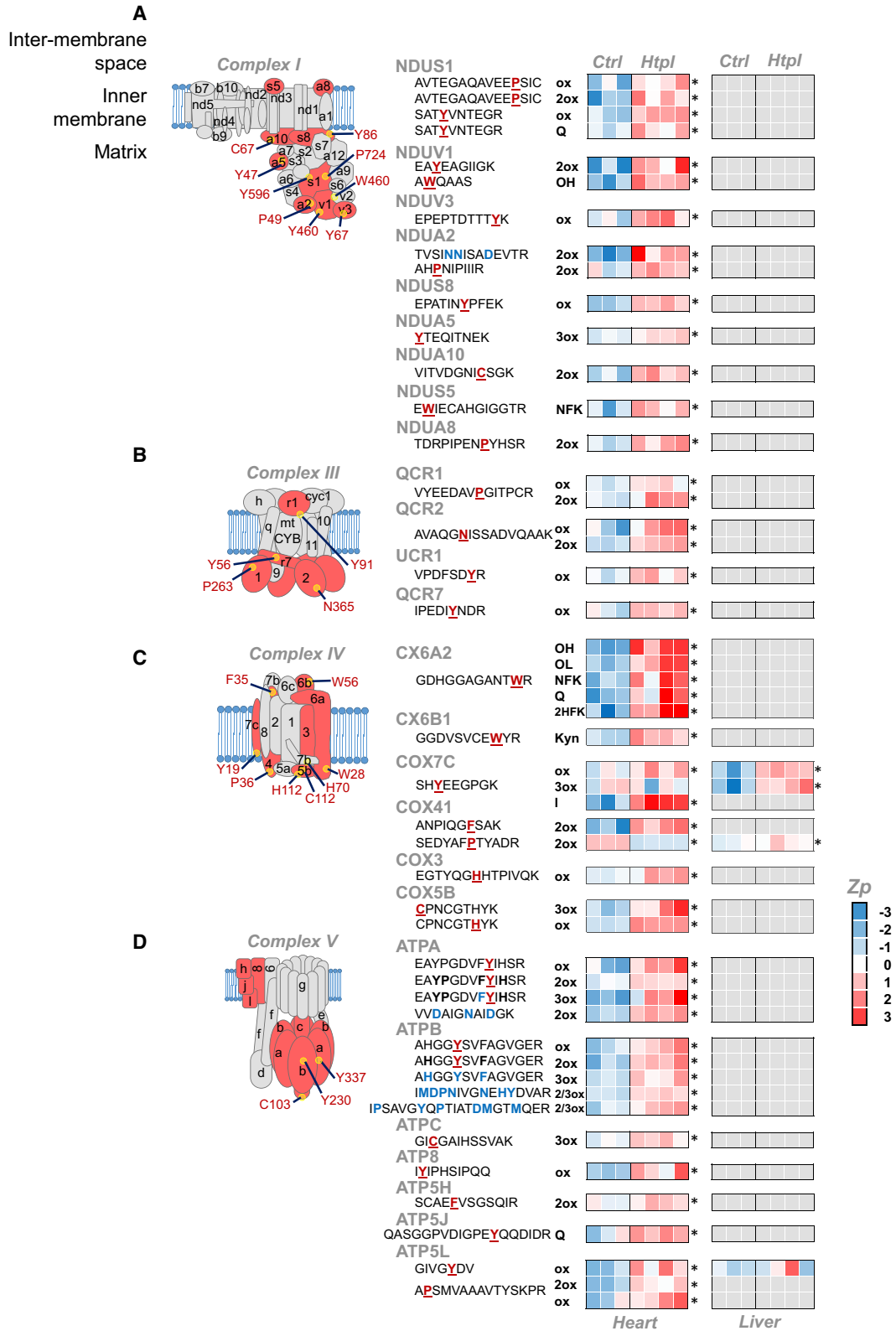
gest that these oxidative modifications compromise OXPHOS protein function. This view is supported by the decreased ATP synthesis capacity we have detected in heteroplasmic heart.

## DISCUSSION

Comprehensive analysis of PTMs in mass spectrometry data obtained through shotgun approaches has historically presented a bioinformatics challenge because conventional (or closed) database searches can only identify predefined subsets of modifications. This limitation impedes truly hypothesis-free analysis of modifications, to the point that the proportion of chemical and PTM in biological systems has remained unknown. The open database search strategy (Chick et al., 2015) provided a starting point for resolving this problem, allowing a truly unbiased high-throughput identification of modified peptides without previous knowledge on the nature of these modifications. An improved fragment-ion indexing method (MSFragger) later demonstrated that OS can be performed on a timescale similar to or shorter than that required for CS (Kong et al., 2017). The open-search strategy has the additional advantage that modified and nonmodified peptide forms are identified simultaneously (Kong et al., 2017). These strategies, however, are not comprehensive and provide limited information on the nature of modified residues. The computational approach presented here enables the open-search strategy to match the performance of CS, identifying the vast majority of detectable peptide modifications. Although Comet-PTM is built on the framework of a well-known database searching engine, it is important to note that the underlying concept can be extrapolated to any other searching engine, opening the way to near-complete analysis of the protein-modification landscape in biological systems using tools already used by the proteomics community. In addition, our method provides information on the most probable site containing the modification with an estimated accuracy of ~85%. It should be noted that, in its current state of development, the method might fail if the peptide contains more than one modification. However, the impact of this limitation on the general performance of the algorithm seems small, since the fraction of multiply modified peptides that were not identified was almost negligible. We also introduced a conservative, three-layered approach to control the FDR of peptide identification. Our local FDR calculation is similar to that used by MSFragger (Kong et al., 2017), with the only difference that we used the decoy/target competition strategy to estimate FDR, whereas Kong et al. modeled local score distributions, in the same 1-Da bins, as a mixture of correct and incorrect identifications and calculated the probability of correct identification by the Bayes rule. As recognized by these authors, their approach did not take into account the parts-per-million levels of accuracy produced by high-resolution

### Figure 5. Functional Distribution of Proteins Post-translationally Modified by Heteroplasmy

(A–C) Functional enrichment analysis of the proteins containing PTMs that are significantly increased (brown scale) or decreased (blue scale) in heart (A), liver (B), or muscle (C) of heteroplasmic mice in relation to control mice. The analysis was performed separately per different kinds of modification (slices). Color intensity represents the number of proteins in each category. Asterisks indicate statistical significance of category enrichment relative to the list of proteins identified in each tissue. Inner circular rings are used for categories at the cellular component level, whereas outer rings correspond to biological processes. (D) Heatmap of the  $Z_{pq}$  values for two modified proteotypic peptides from the two HSP90 isoforms (HSP90AB1 and HSP90AA1) for all control and heteroplasmic replicates in the heart and in liver. These two isoforms are included in the “chaperone protein assembly” functional category displayed in (A) and (B).



**Figure 6. PTMs Induced by Heteroplasmy in OXPHOS Proteins**

(A–D) Scheme of heteroplasmy-induced PTMs in proteins of complex I (A), complex IV (B), complex III (C), and complex V (D) from the OXPHOS system. PTM-containing proteins are highlighted in red, with the spatial position of the modified residue indicated. The heatmaps show statistically significant changes in (legend continued on next page)

instruments (Kong et al., 2017), which we have considered in our model by applying an additional peak FDR estimation. In addition, we found it necessary to introduce a third global FDR filter to avoid identification of low-scoring peptides in local bins or in peaks that, as in the case of oxidation, concentrate high numbers of target PSMs.

The algorithm allowed us to produce peptide maps containing realistic proportions of unmodified peptides, PTMs, and chemical artifacts in mouse tissues, and of amino acid residues subjected to each type of modification. We show that the basal distribution of modifications and modified amino acids is remarkably similar across the three mouse tissues studied, indicating that the algorithm generates reproducible results and that most peptide modifications are of an oxidative nature. Many of the modifications are introduced during sample preparation, including those induced by exposure to Cys-alkylating and amine-directed isobaric labeling reagents; although these agents are thought to be highly specific, in practice they produce secondary reactions with other nucleophiles.

The ability to perform a truly hypothesis-free high-throughput analysis of modifications can contribute to a deeper understanding of the relation between protein chemistry and function; however, the most important biological insight obtainable with shotgun proteomics is how these modifications are altered during normal physiological processes or as a consequence of disease. Detecting these specific alterations allows to concentrate the study on biologically meaningful modifications, obviating the analysis of the abundant artifacts produced by sample preparation. The statistical approach presented here enables quantitative analysis of the modified peptidome within the same framework used to analyze the unmodified proteome; this approach thus allows coherent interpretation of all of the results and opens the way to integrated systems-biology analysis. We demonstrate the performance of our algorithm by simultaneously characterizing, in three tissues of a mouse model, the quantitative impact of heteroplasmy on the proteome and on the modified peptidome. To our knowledge, this is the first study to apply open-search-based approaches together with quantitative MS, allowing comprehensive and unbiased characterization of alterations in the protein modification landscape produced in a biological system.

Our results reveal that the major hallmark of mitochondrial heteroplasmy in 12-week-old animals is oxidative damage to OXPHOS proteins in the heart. This is mainly reflected in the increased levels of oxidative modifications and in the decreased levels of OXPHOS proteins. Our data support the notion that heteroplasmy involving different nonpathological mtDNA variants affects the performance of the OXPHOS system in the heart, resulting in compromised mitochondrial ROS handling that triggers an oxidative stress response and impairs the ability to supply energy to the sarcomere. Heteroplasmy is

known to be related to ROS production (Hämäläinen et al., 2013, 2015) and also affects the levels of contractile proteins in muscle. These alterations do not occur in liver, which is able to resolve heteroplasmy (unpublished data). Our data thus provide a molecular mechanism that explains the functional findings we have observed in the heteroplasmic heart, which include a diminished ability of mitochondria to produce ATP, increased glucose uptake, an exacerbated phosphocreatine/ATP ratio, and mitochondrial disruption and structural alterations.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Mouse Model of Heteroplasmy
  - Generation of Heteroplasmic Mice
  - Mice Breeding
- METHOD DETAILS
  - Benchmarking MS Dataset
  - Preparation of Protein Extracts
  - Protein Digestion, Peptide Labeling, and Fractionation
  - LC-MS Analysis
  - Database Search
  - Comet-PTM
  - SHIFTS
  - Annotation of Modifications
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Peptide Quantification and Statistical Analysis
- DATA AND SOFTWARE AVAILABILITY
  - Code Access

## SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures and four tables and can be found with this article online at <https://doi.org/10.1016/j.celrep.2018.05.080>.

## ACKNOWLEDGMENTS

We thank Simon Bartlett (CNIC) for English editing. This study was supported by competitive grants from the Spanish Ministry of Economy and Competitiveness (MINECO) (BIO2015-67580-P) through the Carlos III Institute of Health-Fondo de Investigación Sanitaria (PRB2, IPT13/0001-ISCI-III-SGFI/FEDER; ProteoRed), by Fundación La Marató TV3, and by FP7-PEOPLE-2013-ITN “Next-Generation Training in Cardiovascular Research and Innovation-Cardionext.” N.B. is a FP7-PEOPLE-2013-ITN-Cardionext Fellow. The CNIC is supported by the MINECO and the Pro-CNIC Foundation, and is a Severo Ochoa Center of Excellence (MINECO Award SEV-2015-0505).

abundance ( $Z_{pq}$ ) of the indicated peptides in heart and liver from control mice (Ctrl) and heteroplasmic mice (Htpl). Gray squares indicate that the modified peptide was not detected in the tissue sample. Modified residues that are unambiguously assigned are underlined bold red letters; tentatively assigned sites are in bold blue letters. Asterisks mark statistically significant changes between control and heteroplasmic samples according to Student's *t* test ( $p < 0.05$ ). The modification type is indicated in the column to the right of the peptide sequences: ox, oxidation; 2ox, dioxidation; 3ox, trioxidation; Kyn, Trp to kynurenine; OH, Trp to hydroxytryptophan; NFK, Trp to N-formyl kynurenine; 2HFK, Trp to 2-hydroxy formyl kynurenine; OL, Trp to oxolactone; Q, Trp to quinone; I, iodination. The conservation scores of these peptides are listed in Table S4, from which the results of Vseq analysis can also be downloaded. See also Table S4.

## AUTHOR CONTRIBUTIONS

N.B., E.B.-K., E.C., and J.V. designed the algorithms. S.M. and M.T.-H. programmed Comet-PTM code. N.B. programmed SHIFTS and analyzed Comet-PTM performance. N.B., E.B.-K., and M.T.-H. programmed the generic integration algorithm and performed the quantitative analysis of PTM. A.V.L.-V., A.L.-P., and J.A.E. developed the heteroplasmic mouse model. E.C., A.V.L.-V., and R.M. prepared the tissue extracts. E.C. and R.M. performed the mass-spectrometry analysis. I.E. and J.M.R. performed the bioinformatic analysis of PTM. N.B., E.B.-K., E.C., and J.V. wrote the manuscript. J.V. directed the research.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 8, 2018

Revised: April 13, 2018

Accepted: May 23, 2018

Published: June 19, 2018

## REFERENCES

- Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., and Ben-Tal, N. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* *44* (W1), W344–W350.
- Ballif, B.A., Villén, J., Beausoleil, S.A., Schwartz, D., and Gygi, S.P. (2004). Phosphoproteomic analysis of the developing mouse brain. *Mol. Cell. Proteomics* *3*, 1093–1101.
- Bern, M., Cai, Y., and Goldberg, D. (2007). Lookup peaks: a hybrid of de novo sequencing and database search for protein identification by tandem mass spectrometry. *Anal. Chem.* *79*, 1393–1400.
- Bonzon-Kulichenko, E., Garcia-Marques, F., Trevisan-Herraz, M., and Vázquez, J. (2015). Revisiting peptide identification by high-accuracy mass spectrometry: problems associated with the use of narrow mass precursor windows. *J. Proteome Res.* *14*, 700–710.
- Chalkley, R.J., Baker, P.R., Medzihradsky, K.F., Lynn, A.J., and Burlingame, A.L. (2008). In-depth analysis of tandem mass spectrometry data from disparate instrument types. *Mol. Cell. Proteomics* *7*, 2386–2398.
- Chen, Y., Chen, W., Cobb, M.H., and Zhao, Y. (2009). PTMap—a sequence alignment software for unrestricted, accurate, and full-spectrum identification of post-translational modification sites. *Proc. Natl. Acad. Sci. USA* *106*, 761–766.
- Chick, J.M., Kolippakkam, D., Nusinow, D.P., Zhai, B., Rad, R., Huttlin, E.L., and Gygi, S.P. (2015). A mass-tolerant database search identifies a large proportion of unassigned spectra in shotgun proteomics as modified peptides. *Nat. Biotechnol.* *33*, 743–749.
- Choi, H., and Nesvizhskii, A.I. (2008). Semisupervised model-based validation of peptide identifications in mass spectrometry-based proteomics. *J. Proteome Res.* *7*, 254–265.
- Cogliati, S., Calvo, E., Loureiro, M., Guaras, A.M., Nieto-Arellano, R., Garcia-Poyatos, C., Ezkurdia, I., Mercader, N., Vázquez, J., and Enriquez, J.A. (2016). Mechanism of super-assembly of respiratory complexes III and IV. *Nature* *539*, 579–582.
- Craven, L., Tuppen, H.A., Greggains, G.D., Harbottle, S.J., Murphy, J.L., Cree, L.M., Murdoch, A.P., Chinnery, P.F., Taylor, R.W., Lightowlers, R.N., et al. (2010). Pronuclear transfer in human embryos to prevent transmission of mitochondrial DNA disease. *Nature* *465*, 82–85.
- Dai, J., Jin, W.H., Sheng, Q.H., Shieh, C.H., Wu, J.R., and Zeng, R. (2007). Protein phosphorylation and expression profiling by Yin-yang multidimensional liquid chromatography (Yin-yang MDLC) mass spectrometry. *J. Proteome Res.* *6*, 250–262.
- Eng, J.K., Jahan, T.A., and Hoopmann, M.R. (2013). Comet: an open-source MS/MS sequence database search tool. *Proteomics* *13*, 22–24.
- Eng, J.K., Hoopmann, M.R., Jahan, T.A., Egerton, J.D., Noble, W.S., and MacCoss, M.J. (2015). A deeper look into Comet—implementation and features. *J. Am. Soc. Mass Spectrom.* *26*, 1865–1874.
- Ferries, S., Perkins, S., Brownridge, P.J., Campbell, A., Evers, P.A., Jones, A.R., and Evers, C.E. (2017). Evaluation of parameters for confident phosphorylation site localization using an orbitrap fusion tribrid mass spectrometer. *J. Proteome Res.* *16*, 3448–3459.
- García-Marqués, F., Trevisan-Herraz, M., Martínez-Martínez, S., Camafeite, E., Jorge, I., Lopez, J.A., Méndez-Barbero, N., Méndez-Ferrer, S., Del Pozo, M.A., Ibáñez, B., et al. (2016). A novel systems-biology algorithm for the analysis of coordinated protein responses using quantitative proteomics. *Mol. Cell. Proteomics* *15*, 1740–1760.
- Griss, J., Perez-Riverol, Y., Lewis, S., Tabb, D.L., Dianes, J.A., Del-Toro, N., Rurik, M., Walzer, M.W., Kohlbacher, O., Hermjakob, H., et al. (2016). Recognizing millions of consistently unidentified spectra across hundreds of shotgun proteomics datasets. *Nat. Methods* *13*, 651–656.
- Hämäläinen, R.H., Ahlqvist, K.J., Ellonen, P., Lepisto, M., Logan, A., Otonkoski, T., Murphy, M.P., and Suomalainen, A. (2015). mtDNA Mutagenesis Disrupts Pluripotent Stem Cell Function by Altering Redox Signaling. *Cell Rep.* *11*, 1614–1624.
- Hämäläinen, R.H., Manninen, T., Koivumäki, H., Kislin, M., Otonkoski, T., and Suomalainen, A. (2013). Tissue- and cell-type-specific manifestations of heteroplasmic mtDNA 3243A>G mutation in human induced pluripotent stem cell-derived disease model. *Proc. Natl. Acad. Sci. USA* *110*, E3622–E3630.
- Hu, C.W., Hsu, C.L., Wang, Y.C., Ishihama, Y., Ku, W.C., Huang, H.C., and Juan, H.F. (2015). Temporal phosphoproteome dynamics induced by an ATP synthase inhibitor citreoviridin. *Mol. Cell. Proteomics* *14*, 3284–3298.
- Huttlin, E.L., Jedrychowski, M.P., Elias, J.E., Goswami, T., Rad, R., Beausoleil, S.A., Villén, J., Haas, W., Sowa, M.E., and Gygi, S.P. (2010). A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell* *143*, 1174–1189.
- Jenuth, J.P., Peterson, A.C., Fu, K., and Shoubridge, E.A. (1996). Random genetic drift in the female germline explains the rapid segregation of mammalian mitochondrial DNA. *Nat. Genet.* *14*, 146–151.
- Jin, W.H., Dai, J., Zhou, H., Xia, Q.C., Zou, H.F., and Zeng, R. (2004). Phosphoproteome analysis of mouse liver using immobilized metal affinity purification and linear ion trap mass spectrometry. *Rapid Commun. Mass Spectrom.* *18*, 2169–2176.
- Keller, A., Nesvizhskii, A.I., Kolker, E., and Aebersold, R. (2002). Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* *74*, 5383–5392.
- Kim, S., and Pevzner, P.A. (2014). MS-GF+ makes progress towards a universal database search tool for proteomics. *Nat. Commun.* *5*, 5277.
- Kong, A.T., Leprevost, F.V., Avtonomov, D.M., Mellacheruvu, D., and Nesvizhskii, A.I. (2017). MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat. Methods* *14*, 513–520.
- Lee, Y., Stiers, K.M., Kain, B.N., and Beamer, L.J. (2014). Compromised catalysis and potential folding defects in vitro studies of missense mutants associated with hereditary phosphoglucomutase 1 deficiency. *J. Biol. Chem.* *289*, 32010–32019.
- Leyfer, D., and Weng, Z. (2005). Genome-wide decoding of hierarchical modular structure of transcriptional regulation by cis-element and expression clustering. *Bioinformatics* *21* (Suppl 2), ii197–ii203.
- Link, A.J., Eng, J., Schieltz, D.M., Carmack, E., Mize, G.J., Morris, D.R., Garvik, B.M., and Yates, J.R., 3rd. (1999). Direct analysis of protein complexes using mass spectrometry. *Nat. Biotechnol.* *17*, 676–682.
- Liu, L., and Pilch, P.F. (2016). PTRF/Cavin-1 promotes efficient ribosomal RNA transcription in response to metabolic challenges. *eLife* *5*, e17508.
- López-Carballo, G., Moreno, L., Masiá, S., Pérez, P., and Baretino, D. (2002). Activation of the phosphatidylinositol 3-kinase/Akt signaling pathway by retinoic acid is required for neural differentiation of SH-SY5Y human neuroblastoma cells. *J. Biol. Chem.* *277*, 25297–25304.

- Lundby, A., Andersen, M.N., Steffensen, A.B., Horn, H., Kelstrup, C.D., Francavilla, C., Jensen, L.J., Schmitt, N., Thomsen, M.B., and Olsen, J.V. (2013). In vivo phosphoproteomics analysis reveals the cardiac targets of  $\beta$ -adrenergic receptor signaling. *Sci. Signal.* **6**, rs11.
- Ma, C.W., and Lam, H. (2014). Hunting for unexpected post-translational modifications by spectral library searching with tier-wise scoring. *J. Proteome Res.* **13**, 2262–2271.
- Martínez-Acedo, P., Núñez, E., Gómez, F.J., Moreno, M., Ramos, E., Izquierdo-Álvarez, A., Miró-Casas, E., Mesa, R., Rodríguez, P., Martínez-Ruiz, A., et al. (2012). A novel strategy for global analysis of the dynamic thiol redox proteome. *Mol. Cell. Proteomics* **11**, 800–813.
- Meier, A., and Söding, J. (2015). Automatic prediction of protein 3D structures by probabilistic multi-template homology modeling. *PLoS Comput. Biol.* **11**, e1004343.
- Mollapour, M., and Neckers, L. (2012). Post-translational modifications of Hsp90 and their contributions to chaperone regulation. *Biochim. Biophys. Acta* **1823**, 648–655.
- Navarro, P., and Vázquez, J. (2009). A refined method to calculate false discovery rates for peptide identification using decoy databases. *J. Proteome Res.* **8**, 1792–1796.
- Navarro, P., Trevisan-Herraz, M., Bonzon-Kulichenko, E., Núñez, E., Martínez-Acedo, P., Pérez-Hernández, D., Jorge, I., Mesa, R., Calvo, E., Carrascal, M., et al. (2014). General statistical framework for quantitative proteomics by stable isotope labeling. *J. Proteome Res.* **13**, 1234–1247.
- Rigbolt, K.T., Prokhorova, T.A., Akimov, V., Henningsen, J., Johansen, P.T., Kratchmarova, I., Kassem, M., Mann, M., Olsen, J.V., and Blagoev, B. (2011). System-wide temporal characterization of the proteome and phosphoproteome of human embryonic stem cell differentiation. *Sci. Signal.* **4**, rs3.
- Sharpley, M.S., Marciniak, C., Eckel-Mahan, K., McManus, M., Crimi, M., Waymire, K., Lin, C.S., Masubuchi, S., Friend, N., Koike, M., et al. (2012). Heteroplasmy of mouse mtDNA is genetically unstable and results in altered behavior and cognition. *Cell* **151**, 333–343.
- Shortreed, M.R., Wenger, C.D., Frey, B.L., Sheynkman, G.M., Scaif, M., Keller, M.P., Attie, A.D., and Smith, L.M. (2015). Global identification of protein post-translational modifications in a single-pass database search. *J. Proteome Res.* **14**, 4714–4720.
- Skinner, O.S., and Kelleher, N.L. (2015). Illuminating the dark matter of shotgun proteomics. *Nat. Biotechnol.* **33**, 717–718.
- van Pée, K.H., and Patallo, E.P. (2006). Flavin-dependent halogenases involved in secondary metabolism in bacteria. *Appl. Microbiol. Biotechnol.* **70**, 631–641.
- Wang, Y.T., Pan, S.H., Tsai, C.F., Kuo, T.C., Hsu, Y.L., Yen, H.Y., Choong, W.K., Wu, H.Y., Liao, Y.C., Hong, T.M., et al. (2017). Phosphoproteomics reveals HMG1, a CK2 substrate, as a drug-resistant target in non-small cell lung cancer. *Sci. Rep.* **7**, 44021.
- Wiśniewski, J.R., Ostasiewicz, P., and Mann, M. (2011). High recovery FASP applied to the proteomic analysis of microdissected formalin fixed paraffin embedded cancer tissues retrieves known colon cancer markers. *J. Proteome Res.* **10**, 3040–3049.
- Wolf, D.P., Mitalipov, N., and Mitalipov, S. (2015). Mitochondrial replacement therapy in reproductive medicine. *Trends Mol. Med.* **21**, 68–76.

## STAR★METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE  | SOURCE  | IDENTIFIER  |
|--|---|---|
| <b>Chemicals</b>   |   |   |
| Filter Aided Sample Preparation (FASP) kit   | Expedeon  | Cat# SKU: 44250   |
| 10 plex-TMT  | Thermo Fisher Scientific  | Cat# 90110  |
| OASIS HLB extraction cartridges  | Waters Corp.  | Cat# 186000383  |
| high pH reversed-phase peptide fractionation kit   | Thermo Fisher Scientific  | Cat# PI84868  |
| EASY-nLC 1200 instrument   | Thermo Fisher Scientific  | Cat# LC120  |
| EASY-Spray NG Ion Source   | Thermo Fisher Scientific  | Cat# ES082  |
| Q Exactive HF mass spectrometer  | Thermo Fisher Scientific  | Cat# IQLAAEGAAPFALGMBFZ   |
| Acclaim PepMap100 C18 100 A nanoViper Trap column.   | Thermo Fisher Scientific  | Cat# 11352013   |
| EASY-Spray column, 50 cm × 75 μm ID, PepMap RSLC C18, 2 μm particles, 100 A pore size  | Thermo Fisher Scientific  | Cat# ES803  |
| <b>Deposited Data</b>  |   |   |
| Raw and analyzed data  | This paper  | <a href="ftp://ftp.peptideatlas.org/">ftp://ftp.peptideatlas.org/</a><br>username: PASS01085 password: TC4334fh.    |
| <b>Experimental Models: Organisms/Strains</b>  |   |   |
| Mouse: 12-week-old BL/6 <sup>C57</sup> , BL/6 <sup>NZB</sup> and BL/6 <sup>C57-NZB</sup> males.                                | José Antonio Enríquez Lab, Centro Nacional de Investigaciones Cardiovasculares (CNIC) | BL/6 <sup>C57</sup> , BL/6 <sup>NZB</sup> and BL/6 <sup>C57-NZB</sup>   |
| <b>Software and algorithms</b>   |   |   |
| Comet release 2016.01  | <a href="#">Eng et al., 2013, 2015</a>  | <a href="http://comet-ms.sourceforge.net">http://comet-ms.sourceforge.net</a>                                       |
| Comet-PTM  | This paper  | <a href="https://github.com/CNIC-Proteomics/Comet-PTM">https://github.com/CNIC-Proteomics/Comet-PTM</a>             |
| SHIFTS (Systematic Hypothesis-free Identification of modifications with controlled FDR based on ultra-Tolerant database Search | This paper  | <a href="https://github.com/CNIC-Proteomics/SHIFTS">https://github.com/CNIC-Proteomics/SHIFTS</a>                   |
| Vseq program   | <a href="#">Cogliati et al., 2016</a>   | available upon request  |
| Generic Integration Algorithm (GIA)  | This paper and <a href="#">García-Marqués et al., 2016</a>                            | <a href="https://github.com/CNIC-Proteomics/PTM-Quant-Stats">https://github.com/CNIC-Proteomics/PTM-Quant-Stats</a> |

### CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests should be directed to and will be fulfilled by the Lead Contact, Jesús Vázquez ([jvazquez@cnic.es](mailto:jvazquez@cnic.es)).

### EXPERIMENTAL MODEL AND SUBJECT DETAILS

#### Mouse Model of Heteroplasmy

All animal procedures conformed to EU Directive 86/609/EEC and Recommendation 2007/526/EC regarding the protection of animals used for experimental and other scientific purposes, enforced in Spanish law under Real Decreto 1201/2005. The mice were fed a standard chow diet (5K67 LabDiet).

#### Generation of Heteroplasmic Mice

Heteroplasmic mice were generated by electro-fusing cytoplasts from conplastic BL/6<sup>NZB</sup> zygotes to recipient C57BL/6J<sup>OlaHsd</sup> (BL/6<sup>C57</sup>) one cell embryos, cultured overnight and transplanted as two-cell embryos into pseudo pregnant Hsd:ICR (CD-1<sup>®</sup>) females to complete development to term as previously described ([Jenuth et al., 1996](#)). To the best of our knowledge, no consensus rule to name heteroplasmic mouse strains exists. Here, we propose the following designation to name heteroplasmic mouse strains: NUCLEAR GENOME-mtCYTOPLASMIC GENOME #1 + CYTOPLASMIC GENOME #2 [i.e., C57BL/6J-mtC57BL/6+NZB, a strain

with the nuclear genome of C57BL/6J and the cytoplasmic (mitochondrial) genome of C57BL/6J and NZB). To simplify we are calling it BL/6<sup>C57-NZB</sup> along this report. The female heteroplasmic offspring (named BL/6<sup>C57-NZB</sup>) were mated with C57BL/6J<sup>OlaHsd</sup> males to prevent nuclear genetic drift in our particular mice lines. Only offspring of the established heteroplasmic mice were used.

### Mice Breeding

Heteroplasmic females (BL/6<sup>C57-NZB</sup>) were outcrossed with males BL/6<sup>C57</sup>. Only females with an initial level of NZB heteroplasmy above 20% were used for colony maintenance.

The mice used in this work were 12-week-old control (C57BL/6J<sup>OlaHsd</sup> strain) and heteroplasmic males (containing more than one mtDNA in the same cytoplasm, C57BL/6 background). The effect of heteroplasmy on the PTMs of the proteome of different tissues was studied on liver, heart, and skeletal muscle (gastrocnemius) samples. In the last two tissues, the heteroplasmy was stable, while the liver was selected as a control tissue since it spontaneously selected one of the alternative variants of mtDNA (manuscript in submission). For each tissue, biological replicates from different control (N = 3) and heteroplasmic mice (N = 4) were analyzed.

## METHOD DETAILS

### Benchmarking MS Dataset

To test the performance of the developed algorithms, we used the publicly available HEK293 dataset (Chick et al., 2015), containing 1.121.149 MS/MS spectra in 24 raw files acquired on a Q-Exactive Orbitrap mass spectrometer. For the bench-marking of site localization through CometPTM we used a synthetic phosphopeptide dataset from the Pride database (dataset identifier PXD007058)(Ferries et al., 2017).

### Preparation of Protein Extracts

Mice were sacrificed by cervical dislocation and liver, heart, and skeletal muscle tissues were extracted. 20 mg of each tissue were homogenized in lysis buffer (10mM Tris-HCL pH7.4, 1 mM EDTA, 0.32 M sucrose, 2% SDS) freshly supplemented with protease and phosphatase inhibitors (Roche) and 50 mM DTT, using a MagNA Lyser instrument (Roche). The lysate was boiled for 5 min and cell debris were removed by centrifugation.

### Protein Digestion, Peptide Labeling, and Fractionation

Proteins were treated with 50 mM iodoacetamide (IAM) and digested with trypsin using the Filter Aided Sample Preparation (FASP) digestion kit (Expedeon) (Wiśniewski et al., 2011) according to manufacturer's instructions. Dried peptides were labeled using 10 plex-TMT reagents according to manufacturer's instructions (Thermo Fisher Scientific), desalted on OASIS HLB extraction cartridges (Waters Corp.) (Leyfer and Weng, 2005), separated into 7 fractions using the high pH reversed-phase peptide fractionation kit (Thermo Fisher Scientific) and dried-down before MS analysis.

### LC-MS Analysis

Each fraction of the labeled peptide samples were analyzed using an Easy nano-flow HPLC system (Thermo Fisher Scientific) coupled via a nanoelectrospray ion source (Thermo Fisher Scientific, Bremen, Germany) to a Q Exactive HF mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). C18-based reverse phase separation was used with a 2-cm trap column and a 50-cm analytical column (EASY column, Thermo). Peptides were loaded in buffer A (0.1% formic acid (v/v)) and eluted with a 240 min linear gradient of buffer B (80% acetonitrile, 0.1% formic acid (v/v)) at 200 nL/min. Mass spectra were acquired in a data-dependent manner, with an automatic switch between MS and MS/MS using a top 15 method. MS spectra were acquired in the Orbitrap analyzer with a mass range of 400–1500 m/z and 60,000 resolution. HCD fragmentation was performed at 27 of normalized collision energy and MS/MS spectra were analyzed at 60,000 resolution in the Orbitrap.

### Database Search

Unless indicated otherwise, all searches were performed using Comet release 2016.01 (Eng et al., 2013, 2015) using trypsin digestion with 1 missed cleavages (unless otherwise specified) and fixed Cys carbamidomethylation (57.021464 Da). For heteroplasmic mice data, TMT labeling at N-terminal end and Lys was also considered as a fixed modification (229.162932 Da). Fragment ion tolerance was 0.02 bin, 0 mass offset. Precursor tolerance type and isotope error were set to 1. Precursor charge range was 2-4, maximum precursor charge 5 and maximum fragment charge 3. Only y- and b-ions were used for scoring.

Closed searches (CS) were performed at 5 ppm precursor ion tolerance, using three dynamic modifications: Met oxidation (15.994915), Asn and Gln deamidation (0.984016) and Ser and Thr phosphorylation (79.966331). False discovery rates (FDR) of peptide identifications were calculated using the refined method (Bonzon-Kulichenko et al., 2015; Navarro and Vázquez, 2009); 1% FDR was used as the default criterion for peptide identification.

Open searches (OS) with Comet and Comet-PTM were performed in the same conditions as CS, except that precursor ion tolerance was set to 500 Da.



### Comet-PTM

Comet.PTM was developed by modifying the open-source database search engine (Eng et al., 2015). For every sequence candidate Comet-PTM calculates the difference between theoretical and experimental precursor mass ( $\Delta$ Mass), and adds up this mass iteratively to each one of the amino acid masses in the peptide sequence, calculating a Xcorr score in each one of the possible modified forms of the peptide (Figure 1B). The selected candidate is the modified peptide form that produces the highest Xcorr. This design allows Comet-PTM to reach the score that would have been obtained by performing a targeted CS with the same modification in the same position. Note that the scores are not exactly identical, since CS uses the theoretical mass of the modification and Comet-PTM estimates it from the difference between the precursor mass and the theoretical mass of the non-modified peptide, and experimental errors on this estimate may affect fragment matching. This effect is, however, small when low ppm precursor mass accuracies are used (Figures S1A and S1B). Comet-PTM has a user-selectable option of scoring also the non-modified peptide sequence (even when  $\Delta$ Mass is different from zero), to take into account labile modifications (Kong et al., 2017).

Comet-PTM was developed to take full advantage of the multi-thread design of Comet. Comet-PTM used less than 4 hr to perform a 500 Da-wide OS of 16 LC-MS runs, containing an average of 44,390 MS/MS spectra each, using a computer cluster with 16 nodes, where each node is built of 2 x Intel Xeon E5-2695v2 at 2.40 GHz and contained 46 threads/124 gigabyte.

### SHIFTS

SHIFTS (Systematic Hypothesis-free Identification of modifications with controlled FDR based on ultra-Tolerant database Search) is a program that identifies peaks in the  $\Delta$ Mass distribution, assigns PSM to peaks and calculates FDR for peptide identification (Figure S6). SHIFTS uses as input the Thermo .raw files and the files obtained from Comet-PTM search.

#### Mass Recalibration

SHIFTS first recalibrates precursor peptide masses independently in each raw file. This was done by selecting a population of non-modified peptides with a very high score (user selectable; recommended values are those yielding 0.1% global FDR or lower), which are assumed to be true identifications and are used to calculate the systematic mass error (median deviation in m/z scale), which is assumed to be constant in each raw file. From these data, SHIFTS also calculates the standard deviation of the mass error ( $\sigma_M$ ) using the median absolute deviation (MAD) method.

#### Peak Identification

Recalibrated  $\Delta$ Mass values were binned using 0.001 Da bins to construct the  $\Delta$ Mass distribution. The distribution was smoothed using the median of a 7-point sliding window and then peak apexes were detected as downward zero-crossings in the first derivative of the smoothed curve. Peak widths were similarly calculated as the zero-crossing points of the second derivative; in the current version of SHIFTS they are computed only for informative purposes.

#### Peak Assignment

By default SHIFTS assigns a PSM to the closest  $\Delta$ Mass peak if the mass deviation of the PSM from the peak falls within  $3\sigma_M$ , so that approximately 99% of PSM in each peak are assigned. This value can be user adjusted. PSM not assigned to peaks were considered as orphan PSM.

#### FDR Calculation

SHIFTS calculates FDR of identification using a conventional target/decoy strategy using the corrected Xcorr score (cXcorr) (Choi and Nesvizhskii, 2008; Keller et al., 2002). A global FDR was calculated for each PSM as the ratio of the number of decoy PSMs to the number of target PSMs having a cXcorr equal or higher. Decoy peptides matched by Comet-PTM were observed to be almost as abundant as target peptides in the negative  $\Delta$ Mass region below the peak corresponding to neutral loss of Gly (Figure 2A), where  $\Delta$ Mass peaks were mostly produced by neutral loss of amino acids. For this reason, the global FDR was only calculated in the  $\Delta$ Mass region above  $-56$  Da (Figure 2A). All the PSM are required to have FDR lower than the global FDR, without exception.

In addition, local FDR filters are also applied. Some  $\Delta$ Mass peaks were observed to contain an unusually high number of decoy PSM; to avoid matching false positive target PSM in these peaks, SHIFTS also calculates a peak FDR counting up the number of decoys and target PSM assigned to each peak, and these PSM are required to pass the peak FDR filter in addition to the global FDR filter (Figure 2C). Note that peak FDRs are often very low suggesting that the majority of PSM in these peaks are true, even when they have a low cXcorr. This happens because the probability of finding a decoy PSM in a peak by chance alone is extremely low. SHIFTS avoids matching these low scoring target PSM by applying the global FDR filter.

To apply a local filter to PSM which are not assigned to  $\Delta$ Mass peaks, e.g., to orphan PSM, SHIFTS models the periodic mass distribution of decoy PSM into  $\sim 1$  Da-bins centered at the regions where  $\Delta$ Mass values concentrate, and calculates a local FDR by counting up decoy and target PSM in each one of these regions (Figure 2B). The local FDR filter is applied to orphan PSM in addition to the global FDR filter. Default values for peptide identification were 1% for peak and local FDR; 5% for global FDR. Among the peptides significantly changed by heteroplasmy and suspected to have relevant biological activity, including all the peptides in Table S3, only those that passed analysis with Vseq program (Cogliati et al., 2016) were considered as trustable identifications.

#### Isotopic Correction

SHIFTS also performs a simple isotopic correction to minimize misassignments of the correct monoisotopic peak of the precursor. When two PSM having the same sequence are encountered having a  $\Delta$ Mass difference within 1 ppm of the mass difference expected for either one or two  $^{13}\text{C}$  or one  $^{34}\text{S}$ , the  $\Delta$ Mass of the heaviest precursor is substituted by that of the lightest one.

### Annotation of Modifications

A Python in-house script was used for semisupervised annotation of the nature of peptide modifications. The script searched  $\Delta$ Mass values against Unimod database, taking into account the amino acid modified according to Comet-PTM output and also the preceding and consecutive residues, comparing them with the list of amino acids that could be subjected to the modification according to Unimod. If no amino acid was matched, the modification was considered as unassigned. Residues containing modifications that were fixed in the database search were considered with and without the fixed modification; when not indicated, the amino acid contains the fixed modification (e.g., C\_oxidation means oxidation of carbamidomethylated Cys and K\_oxidation, oxidation of TMT-labeled Lys).  $\Delta$ Mass values that could not be matched were tentatively tested assuming one  $^{13}\text{C}$  misassignment of the monoisotopic mass of the precursor and also as combinations of two modifications from a list of the most abundant modifications found in the corresponding proteome. Unexplained  $\Delta$ Mass values were termed as unknown. MS/MS fragmentation spectra from the most abundant modifications that changed their abundance in heteroplasmic mice and all the peptides in Table S3 were revised using Vseq program (Cogliati et al., 2016).

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Peptide Quantification and Statistical Analysis

The quantitative information from TMT reporter intensities was integrated from the spectrum level to the peptide level and then to the protein level on the basis of the WSPP model (Martínez-Acedo et al., 2012; Navarro et al., 2014) using the Generic Integration Algorithm (GIA) (García-Marqués et al., 2016). The algorithm was modified to include into the statistical model the quantitative values of modified peptides as part of the automated workflow (Figure 3A). Briefly, for each sample  $i$  the values  $x_{qps} = \log_2 S_i / C$  were calculated, where  $S_i$  is the intensity of the TMT reporter corresponding to sample  $i$  in the MS/MS spectrum  $s$  coming from peptide  $p$  and protein  $q$ , and  $C$  is the average intensity of all the TMT reporters from the control samples, which is used as a common reference. The  $\log_2$ -ratio of each peptide ( $x_{qp}$ ) was calculated as the weighted average of its spectra, the protein values ( $x_q$ ) were the weighted average of its peptides, and the grand mean ( $\bar{x}$ ) was calculated as the weighted average of all the protein values (Navarro et al., 2014). The statistical weights of spectra, peptides, and proteins ( $w_{qps}$ ,  $w_{qp}$  and  $w_q$ , respectively) and the variances at each one of the three levels ( $\sigma_S^2$ ,  $\sigma_p^2$ , and  $\sigma_q^2$ , respectively), were calculated as described (Navarro et al., 2014).

The spectrum, peptide, and protein variances and the protein values were first determined including only non-modified peptides (Figure 3B). In a second step, the modified peptides were included in the analysis, which was performed using the variances and protein values calculated previously. For each modified peptide, the standardized variable ( $z_{pq}$ ) was calculated as

$$z_{pq} = (x_{qp} - x_q) \sqrt{w_{pq}} \sqrt{\frac{n_p}{n_p - 1}}, n_p > 1$$

where  $n_p$  is the number of non-modified peptides with which the corresponding protein was quantified.  $z_{pq}$  expresses the deviation between the peptide  $\log_2$ -ratio and the corresponding protein value in units of standard deviation. In absence of changes, the distributions of  $z_{pq}$  followed very closely the normal distribution  $N(0, 1)$  (Figure 3C), validating the accuracy of the model.

Significant abundance changes of modified peptides were detected by Student's  $t$  test comparing the  $z_{pq}$  values of samples from heteroplasmic mice ( $N = 4$ ) with those of control mice ( $N = 3$ ) (Table S2).

## DATA AND SOFTWARE AVAILABILITY

### Code Access

The software needed to execute the whole Comet-PTM pipeline can be downloaded from: <https://github.com/CNIC-Proteomics/Comet-PTM>.

The software for Comet-PTM FDR control can be downloaded from: <https://github.com/CNIC-Proteomics/SHIFTS>.

The software for statistical analysis of quantitative data can be downloaded from <https://github.com/CNIC-Proteomics/PTM-Quant-Stats>. A readme.txt file is provided with basic instructions to install and execute the package. Vseq is available upon request.

The dataset of the raw files, protein databases, search parameters and results reported in this paper is available in the PeptideAtlas repository at <ftp://ftp.peptideatlas.org/> (username: PASS01085, password: TC4334fh).