Quantitative proteomics reveals key roles for post-transcriptional gene regulation in the molecular pathology of FSHD.

Sujatha Jagannathan[1,2,3,#,*], Yuko Ogata[4,$], Philip R. Gafken[4], Stephen J. Tapscott[3*] and Robert K. Bradley[1,2*]

[1]Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA

[2]Computational Biology Program, Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA

[3]Human Biology Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA

[4]Proteomics Facility, Fred Hutchinson Cancer Research Center

Current affiliation: Department of Biochemistry and Molecular Genetics & RNA Bioscience Initiative, University of Colorado Denver School of Medicine, Aurora, CO, 80045, USA.

$ Current affiliation: Just Biotherapeutics in Seattle, WA 98109

*Corresponding authors

Email: rbradley@fredhutch.org; sujatha.jagannathan@ucdenver.edu; stapscot@fredhutch.org

Abstract

DUX4 is a transcription factor whose misexpression in skeletal muscle causes facioscapulohumeral muscular dystrophy (FSHD). While DUX4's transcriptional activity has been extensively characterized, the DUX4-induced proteome remains undescribed. Here, we report concurrent measurement of RNA and protein levels in DUX4-expressing cells via RNA-seq and quantitative mass spectrometry. DUX4 transcriptional targets were robustly translated, confirming the likely clinical relevance of proposed FSHD biomarkers. However, a multitude of mRNAs and proteins exhibited discordant expression changes upon DUX4 expression. Our dataset revealed unexpected proteomic, but not transcriptomic, dysregulation of diverse molecular pathways, including Golgi apparatus fragmentation, as well as extensive post-transcriptional buffering of stress response genes. Key components of RNA degradation machinery, including UPF1, UPF3B, and XRN1, exhibited suppressed protein, but not mRNA, levels, explaining the build-up of aberrant RNAs that characterizes DUX4-expressing cells. Our results provide a resource for the FSHD community and illustrate the importance of post-transcriptional process to DUX4-induced pathology.

Introduction

Facioscapulohumeral muscular dystrophy is caused by the inappropriate expression of an early embryonic transcriptional activator, DUX4, in adult muscle, leading to cell death (1, 2). Decades of work have generated a detailed parts-list of the genes and pathways affected by DUX4 that may underlie FSHD pathophysiology (3-10). Yet, an integrated model for how those DUX4-induced changes lead to disease has remained elusive (11-13). Considering that transient and pulsatile expression of DUX4 is sufficient induce pathology and cell death (14), it is particularly important that we understand the cellular events and pathways set in motion by DUX4 that lead to eventual cell death in order to develop effective therapeutics for FSHD.

DUX4 induces changes in the expression of hundreds of genes that impact dozens of highly interconnected pathways (3-10), making a cause-effect relationship between the measured gene expression and the observed pathology difficult to discern. Given that DUX4 is a strong transcription factor, most studies on DUX4 activity have focused on measuring gene expression at the transcript-level (3, 14, 15), making the assumption that the transcriptome is an accurate representation of the cellular proteome in DUX4-expressing cells. While this is a reasonable assumption, it is well known that RNA and protein levels are not always concordant and post-transcriptional regulation can cause them to significantly diverge (16). The few early proteomics studies that exist were conducted on FSHD muscle biopsies and lack the depth necessary to draw meaningful comparisons with the DUX4 transcriptome (17-19). Given our recent discovery that DUX4 induces proteolysis of a key RNA binding protein, UPF1 (7), we hypothesized that post-transcriptional gene regulation could be an important factor to consider in understanding DUX4 biology. Hence, we set out to generate reliable protein-level measurements of DUX4-induced gene expression and thereby elucidate the extent of post-transcriptional gene regulation in DUX4-expressing cells.

Using SILAC-based quantitative mass spectrometry (20), we measured the fold-change at the protein level for ~4000 genes upon DUX4 expression, with high confidence. Comparing the transcript fold-change to the protein fold-change shows three distinct patterns of expression for different subsets of genes: 1) Concordant change in expression at the RNA and protein level for a number of direct DUX4 transcriptional targets; 2) Post-transcriptional buffering of the expression of a large number of genes, especially of those involved in stress response; and 3) Discordant gene expression change at the RNA versus protein level that included genes involved in RNA surveillance. Together, these findings highlight the importance of considering the expressed proteome to fully understand DUX4 biology and the FSHD disease process.

Results

Determining protein fold-change in DUX4-expressing cells via quantitative mass spectrometry.

In order to measure DUX4-induced changes to the cellular proteome, we conducted SILAC-based mass spectrometry in two independent DUX4 expression systems (**Figure 1A).** We have previously shown that DUX4 expressed via a lentiviral vector versus an inducible transgene integrated into the genome of a myoblast cell line both yield comparable gene expression profiles (21). Here, we use both of the expression systems to provide a stringent biological replicate for our proteomic analysis.

In a pilot-scale proteomics experiment, human muscle cells that were adapted to light or heavy SILAC media for 3 weeks were transduced with lentivirus carrying DUX4 (vDUX4) or GFP (vGFP) expression constructs and samples were collected 24 and 36 hours post-transduction. In an independent experiment, cells carrying a doxycycline-inducible DUX4 transgene (iDUX4; (21)) were adapted to SILAC media for 3 weeks and DUX4 expression was induced with 1µg/ml of doxycycline for 14 hours in two replicates carrying heavy and light SILAC labels. Paired controls with no treatment were also collected with both heavy and light labels. Total protein from cells expressing DUX4 were mixed with an equal amount of total protein from cells without DUX4 expression containing the opposite SILAC label to generate samples that were then subjected to mass spectrometry.

Peptide-spectrum matches (PSMs) with quantified heavy to light ratios were subject to thorough screening for quality (e.g., filtering out single-peak spectra and spectra without unique mapping; see *Materials & Methods* for further details; Supplemental Data Set 1). Peptides mapping to a DUX4 target gene, ZSCAN4, from both vDUX4 and iDUX4 datasets showed highly skewed $\log_2$ (DUX4/Control) ratio that showed significant upregulation of the proteins upon DUX4

97 expression (**Figure 1B-C**). In contrast, plotting the $\log_2$ (DUX4/Control) ratio of all individual
98 peptides mapping to a housekeeping gene RPL15 showed that the ratio is centered around zero
99 (**Figure 1B-C**), as would be expected for a gene with no differential expression upon DUX4
100 induction. These example plots illustrate the strong agreement between the expected and
101 observed protein fold-change values determined by SILAC mass spectrometry. Moreover, out of
102 the 65 genes identified by Yao *et al* (22) as potential FSHD biomarkers based on transcriptome
103 analysis of FSHD patient samples, 8 were quantified in the vDUX4 proteomics study and 25 were
104 quantified in the iDUX4 proteomics study and both show high induction at the protein level (**Figure**
105 **1D, E**). Note that the lower number of peptides (and hence quantified proteins) from the vDUX4
106 sample indicates the lower depth of this dataset, and yet yields fold-changes that are highly
107 consistent with the iDUX4 dataset with a higher depth.

108

109 **Assessing concordance of RNA and protein expression fold-change.**

110 Next, using the iDUX4 dataset, we performed peptide to protein summarization via median
111 heavy/light ratios of all the peptides mapping to a certain protein in both replicates to obtain gene-
112 level $\log_2$ (DUX4/Control) ratios (**Figure S1**; Supplemental Data Set 2). After filtering out genes
113 that were only observed in one of the two label-swap replicates, we obtained quantitative
114 proteomics information for 4005 genes, 3961 of which also had a corresponding RNA-seq
115 measurement (**Figure 2A;** RNA-seq data previously reported in (21)). The lower number of genes
116 quantified via proteomics compared to RNA-seq is expected as proteomics is known to have lower
117 sensitivity compared to RNA-seq.

118 To qualitatively compare the RNA and protein expression level changes upon DUX4
119 expression, we assessed the overlap of genes with an expression change of 4-fold or above.
120 Among genes that are upregulated (> 2 log2 fold change), the concordance between RNA and
121 protein was roughly 40-50%, whereas similarly downregulated genes show very little concordance

(**Figure 2A**). To obtain a more quantitative measure of concordance, we generated a scatter plot of the RNA versus protein fold-change for the 3961 genes (**Figure 2B**). We found a reasonable level of correlation between these values with a Pearson's correlation coefficient, r, of 0.51 (p-value < 2.2e-16).

To assess whether similar pathways were affected at the RNA versus protein level, we conducted gene ontology analysis for genes up- or down-regulated at the RNA and protein (**Figure 2C**). Surprisingly, we observed that the pathways that are affected at the RNA versus protein levels are quite distinct. While the transcript level changes occur in genes involved in transcription and mRNA processing, protein-level changes impact pathways including humoral immune response, proteolysis and exocytosis. As the exocytosis pathway has not been implicated in any of the previous DUX4 gene expression studies, we sought to examine this phenomenon further by imaging the Golgi apparatus, which is the source of exocytotic vesicles in the cell (23). We found that DUX4 expressing cells showed severe fragmentation of Golgi apparatus, which could be an indicator of a perturbation in the cellular secretory pathways (24) (**Figure 2D;** quantified in **Figure 2E**).

Taken together, these results demonstrate that analyzing the protein measurements may give us insights that are not discernable in the transcriptome fold-change analysis performed in earlier studies.

Post-transcriptional buffering of stress response genes may exacerbate DUX4 toxicity.

Though many of the genes induced at the transcript level are largely also induced at the protein-level, a subset of genes showed no change in their protein level despite their transcripts were up- or down-regulated to a significant degree (678 genes, shaded blue in **Figure 3A**), indicating post-transcriptional buffering of the protein levels. Most notably, several housekeeping genes that respond to protein folding stress or dsRNA-induced stress showed transcriptional upregulation with minimal protein-level upregulation (**Figure 3B**).

Given that both unfolded protein and dsRNA induced stresses converge in phosphorylation of eIF2α and lead to translation inhibition (25), we asked if the timing of transcription of various stress-response genes coincide with translational downregulation. We found that HSPA5, a prominent marker of the unfolded protein response pathway (26), shows transcriptional upregulation at a time period that temporally coincides with eIF2α phosphorylation and reduced incorporation of S35-labeled methionine, a proxy for bulk translation efficiency (**Figure 3 C-E**). These data demonstrate that translation inhibition caused by various cellular stresses and the resulting post-transcriptional buffering prevents DUX4-expressing cells from mounting a robust stress response.

Post-transcriptional modulation of RNA quality control pathway by DUX4.

Next, we focused our analysis on the subset of genes that showed significant changes at the protein level with either no change in their transcript abundance or a change in the opposite direction (198 genes shown as 'gold' circles in **Figure 4A**). Pathway analysis did not reveal any significant trends among these genes. So instead, we decided to focus on one of the pathways that we have previously shown to be post-transcriptionally modulated – namely, the nonsense-mediated RNA decay (NMD) pathway (7).

A diagram showing RNA- vs protein-level changes to various components of this pathway demonstrates substantial post-transcriptional regulation in this pathway (**Figure 4B**). Many of these genes, including UPF1, UPF2, UPF3B and XRN1, showed downregulation at the protein level. The downregulation of XRN1 is of particular interest as it is the 5'-3' exonuclease that degrades NMD targets upon cleavage by the endonuclease, SMG6 (27). Moreover, SMG6 too is downregulated to a log2 fold change of -4.7, though it is only detected via a single peptide and hence was filtered out of our analysis. Thus, DUX4-induced NMD inhibition appears to be a result

of post-transcriptional downregulation of multiple key players of the NMD pathway, which explains the severity of NMD inhibition in DUX4-expressing cells.

Post-transcriptional downregulation of a gene can be achieved via two means: reduced translation or increased protein degradation. We have previously shown that DUX4 induces proteasome-mediated degradation of UPF1 (7). Hence, we asked if DUX4 affects known components and regulators of the ubiquitin proteasome. A scatterplot of all ubiquitin proteasome regulators shows a change in the expression of several such genes, one or more of which may underlie the rapid degradation of UPF1 (**Figure 4C**). Further studies are needed to understand the precise molecular mechanism behind this regulatory pathway and its downstream consequences. In summary, we propose that the post-transcriptional gene regulation plays a critical role in inhibiting NMD and perturbing the proteostasis in DUX4-expressing cells and may thus underlie key aspects of FSHD pathology (**Figure 4D**).

Finally, in order to enable researchers and patients in the FSHD community to access the data generated in this study, we developed a web tool for easy visualization of these data (Screenshot shown in **Figure S2**). This tool can be freely accessed at the following URL: https://dynamicrna.shinyapps.io/dataviz/.

Discussion

Most of the highly induced DUX4 transcriptional targets are germline and early embryonic genes that are normally never expressed in adult muscle (3). So, it is possible that despite being expressed at the transcript level, such genes may be translated poorly and/or be degraded rapidly upon translation due to the lack of cell-type chaperones or other factors. Here, we show using quantitative mass spectrometry that DUX4-induced transcripts are efficiently translated to produce stable proteins in the muscle cell. This is an important confirmation for the altered identity of DUX4-expressing cells, and also provides validation to pursue some of these proteins as potential FSHD biomarkers (22).

Next, we asked if the changes in the DUX4 proteome are largely reflective of the changes to the transcriptome. We find that this is not the case. While there is a positive correlation between these measurements (Pearson's correlation coefficient 0.51), hundreds of genes deviate from this trend. GO analysis of the most differentially expressed genes at the transcript versus protein level shows splicing and RNA processing as being the prominent categories impacted at the transcript level while protein-level changes impact an entirely different set of pathways. We take these results as indication that the transcriptome analysis does not paint a complete picture of DUX4 biology and needs to be complemented with proteome analysis to develop a more thorough understanding of how DUX4 misexpression causes FSHD.

We next pursued the various mechanisms by which protein-level changes deviate from the corresponding transcriptomic changes. We found that genes induced by dsRNA and unfolded protein stress are transcriptionally induced, but translationally buffered as a result of the translation repression that accompanies these stress-response pathways. As a result, we postulate that the DUX4-expressing cells are unable to mount a robust stress response, despite inducing the transcripts necessary to alleviate stress. We also show that multiple proteins in the RNA surveillance pathway, including UPF1 and XRN1, are downregulated at the protein level,

which may explain the drastic reduction in RNA quality control capacity in DUX4-expressing cells. From the proteomics data, we have identified a number of genes involved in the ubiquitin proteasome pathway that could impact the stability of proteins, which should serve as a starting point for further investigation into this novel regulatory mechanism of RNA surveillance in DUX4-expressing cells.

Materials and Methods

Accession codes. Mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository (Vizcaino et al., 2013) with the dataset identifier PXD010221 (Reviewer account details: Username: reviewer37110@ebi.ac.uk; Password: aMrE4BW0). The RNA-seq data are available through the NCBI SRA database under accession number GSE85461 (21).

Cell culture and SILAC labeling. Proliferating human myoblasts were cultured in F10 medium (Gibco/Life Technologies) supplemented with 20% fetal bovine serum (Thermo Scientific), 10ng bFGF (Life Technologies), 1μM dexamethasone (Sigma) and 50U/50μg penicillin/streptomycin (Life Technologies). To induce DUX4 expression, 1μg/ml of doxycycline was added for 8 or 14 hours, as indicated. Cells were labeled in SILAC media containing heavy lysine (Lys8) and arginine (Arg10) for 3 weeks before DUX4 induction experiments.

Gel slice digestion. Total RNA and protein were extracted from whole cells using TRIzol^{RT} reagent (Ambion) following the manufacturer's instructions. 50 μg of total protein was subjected to SDS PAGE using a 4-15% bis-TRIS gel. The gel was stained using GelCode blue (Pierce) as per manufacturer's instructions, destained overnight in ultrapure water and the entire lane containing protein was cut into 16 fractions using a GelCutter (Gel company Inc.). Individual gel slices in 1.5 mL tubes (Eppendorf) were consecutively washed with water and incubated with 25mM ammonium bicarbonate, 50% acetonitrile for 2 hrs. The gel pieces were dehydrated with acetonitrile. And the dried gel slices were reduced by covering them with 10 mM dithiothreitol in 100 mM ammonium bicarbonate and heating them at 56^{o}C for 45 min. The solution was removed and discarded. The gel slices were alkylated by covering them with a solution of 50 mM iodoacetamide in 100 mM ammonium bicarbonate and incubating in the dark at ambient

246 temperature for 30 min. The solution was removed and discarded. The gel slices were dehydrated
247 with acetonitrile, then washed with 100 mM ammonium bicarbonate 100 mM for 10 min. The
248 solution was removed, discarded and the gel slices were dehydrated once again with acetonitrile.
249 After removing acetonitrile, the gel slices were then hydrated with 5 ng/uL sequencing grade
250 trypsin (Promega) in 50 mM ammonium bicarbonate and digested overnight at 37°C on an orbital
251 shaker. Following digestion, the supernatants were collected, and the gel slices were washed with
252 0.1% trifluoroacetic acid, and after 30 min an equal volume of acetonitrile was added followed by
253 washing for an additional 1 hour. The original digestion supernatant and the wash for a single
254 sample were combined into a single tube and dried by vacuum centrifugation. The digestion
255 products were desalted using Ziptips (Millipore) per the manufacturer's instructions, eluted with
256 70% acetonitrile/0.1% trifluoroacetic acid, and dried by vacuum centrifugation.

257

258 **Mass spectrometry.** The desalted material was resuspended in 20 µL of 2% acetonitrile in 0.1%
259 formic acid, and 18 µL was analyzed using one of two LC/ESI MS/MS configurations. The first
260 configuration consisted of an Easy-nLC II (Thermo Scientific) coupled to a Orbitrap Elite ETD
261 (Thermo Scientific) mass spectrometer using a trap-column configuration as described (28). A
262 trap of 100 µm × 20 mm packed with Magic C_{18}AQ (5-µm, 200 Å resin; Michrom Bioresources)
263 packing material was used for in-line desalting and a column of 75 µm × 250 mm packed with
264 C_{18}AQ (5-µm, 100 Å resin; Michrom Bioresources) was used for analytical peptide separations.
265 Chromatographic separations were carried out using a 60-minute gradient from 5% to 35%
266 solvent B (solvent A: 0.1% formic acid, solvent B: 0.1% formic acid in acetonitrile) at a flowrate of
267 300 nL/min. The analytical column temperature was maintained at 40℃. The Orbitrap Elite
268 instrument was operated in the data-dependent mode, switching automatically between MS
269 survey scans in the Orbitrap (AGC target value 1E6, resolution 240,000, and maximum injection
270 time 250 ms) and collision induced dissociation (CID) MS/MS spectra acquisition in the linear ion

trap (AGC target value of 10,000 and injection time 100 ms). The twenty most intense precursor ions from the OrbiTrap full scan were each consecutively selected for fragmentation by CID in the linear ion trap using a normalized collision energy of 35%. Ions of +2 and +3 charge states were selected for MS/MS and selected ions were dynamically excluded for 30 seconds. The second configuration consisted of an Easy nanoLC 1000 (Thermo Scientific) HPLC connected to an OrbiTrap Fusion (Thermo Scientific) mass spectrometer. In-line chromatographic separations (no trap column) were carried out using a 75 µm × 400 mm column packed with Magic C_{18}AQ (5-µm, 100 Å resin; Michrom Bioresources) packing material at a flowrate of 300 nL/min. Chromatographic elution consisted of a 90-minute gradient from 3% to 27% B and the column temperature was maintained at 40℃. The OrbiTrap Fusion was operated in the 2 second "top speed" data dependent acquisition mode with MS survey scans in the OrbiTrap at least every 2 seconds (AGC target value 4E5, resolution 120,000, and maximum injection time of 50 ms). Quadrupole isolation was set to 1.6 FWHM and higher energy collision dissociation (HCD) was used for fragmentation at a collision energy of 28% and MS/MS detection was carried out in the linear ion trap set at rapid scan speed (injection time of 250 ms and AGC target of 10E2). Positively charged ions from 2 to 6 were selected for MS/MS and selected ions were dynamically excluded for 30 seconds.

Data analysis and statistical methods. Qualitative and quantitative data analysis were performed using Proteome Discoverer 2.1 (Thermo Scientific). The data were searched against a human UniProt database (downloaded 11-04-16) that was appended with protein sequences from the common repository of adventitious proteins (cRAP; www.thegpm.org/crap/) an *in silico* translation products of noncanonical transcript isoforms stabilized due to NMD inhibition. In downstream analyses, peptides that only mapped to NMD targets were not considered any further in the current study and will be pursued in a future investigation. Searches were conducted with

296 the trypsin enzyme specificity. The precursor ion tolerance was set to 10 ppm and the fragment
297 ion tolerance was set to 0.6 Da. Variable modifications were set for oxidation on methionine
298 (+15.995 Da), carbamidomethyl (+57.021 Da) on cysteine, and acetylation (+42.010 Da) on the
299 N-terminus of proteins. Heavy SILAC amino acids for lysine (+8.014 Da) and arginine (10.008)
300 were also accounted for in the analysis as variable modifications. All search results were
301 evaluated by Percolator (29) for false discovery rate (FDR) evaluation of the identified peptides.
302 Peptide identifications were filtered to a peptide FDR of 1%. All downstream data analysis was
303 conducted using the R statistical programming language. The complete code to reproduce the
304 analyses and figures in this manuscript is deposited in github, available at the following URL:
305 https://github.com/sjaganna/2018-jagannathan_et_al.
306

307 **Gene Ontology analysis.** GO analysis was performed via the Overrepresentation Enrichment
308 Analysis method using WebGestalt server (pmid: 15980575; www.webgestalt.org). One of the
309 GO categories identified as enriched in this analysis was skin development, which we
310 subsequently removed from Figure 2C as many of the genes that contributed to this GO category
311 were extracellular proteins including keratins that could be environmental contaminants.
312

313 **Immunofluorescence microscopy.** Cells were permeabilized with PBS containing 0.1% Triton
314 X-100 for 5 minutes at room temperature and rinsed thrice in PBS. Primary antibody against
315 GM130 (Bethyl Laboratories, Cat # A303-402A-T) was diluted 1:200 in PBS and incubated for 1
316 hour at room temperature. After three washes in PBS, secondary anti-Rabbit TRITC (Jackson
317 ImmunoResearch, cat # 711-025-152) diluted 1:400 was added and incubated for 45 minutes at
318 room temperature. Cells were washed thrice in PBS with the nuclear counterstain DAPI included
319 in the final wash. Images were collected on a Cytation 5 multimode reader (BioTek) and analyzed
320 using GenPrime software (BioTek).

Figure legends

Figure 1. Quantitative mass spectrometry of DUX4-expressing cells.

A) Schematic of experimental set up to measure protein fold-change in cells expressing vDUX4 or iDUX4.

B-C) Histogram shows the DUX4/Control ratios (log2) for all peptides for the DUX4 target gene, ZSCAN4 and the housekeeping gene, RPL15 in the vDUX4 (B) and iDUX4 (C) datasets.

D-E) FSHD biomarker transcripts identified by Yao et al (2015) in vDUX4 (D) and iDUX4 datasets (E).

Figure 2. Concordant changes in RNA and protein abundance of several DUX4 transcriptional targets.

A) Table showing overlap of genes with a quantified RNA and protein level change in expression.

B) Scatter plot of RNA versus protein log2 fold-change (DUX4/control) of all genes.

C) GO analysis of genes that are up- or down-regulated > 4-fold at the RNA level or protein level. Multiple testing correction done via Benjamini-Hochberg procedure.

D) Fluorescence micrographs of cells with and without DUX4 induction for 14h, stained with DAPI, anti-GM130 (Golgi) and phalloidin-555 (actin).

E) Quantification of percent of cells with DUX4 positive nuclei and fragmented Golgi apparatus.

Figure 3. Extensive post-transcriptional buffering of stress response genes in DUX4-expressing cells.

A) Scatter plot of RNA versus protein log2 fold-change (DUX4/control).

B) Representative genes that show post-transcriptional buffering.

C) qRT-PCR for a chaperone HSPA5 over a time course of vDUX4 expression.

D) Western blot showing phosphorylation of eIF2a over DUX4 expression time course.

E) S35-methionine incorporation over DUX4 time course.

Figure 4. DUX4 induces post-transcriptional gene regulation.

A) Scatter plot of RNA versus protein log2 fold-change (DUX4/control), highlighting genes that are post-transcriptionally regulated.

B) Schematic representation of RNA- and protein-level changes for the genes involved in mRNA surveillance. Colors represent a heat map of actual fold-change values. Proteins outlined by dotted lines represent those lacking high-quality data.

C) Scatter plot of RNA versus protein log2 fold-change (DUX4/control) of genes in the ubiquitin proteasome pathway.

D) Model for DUX4-induced post-transcriptional gene regulation.

359 **Supplementary Figures.**
360
361
362 **Figure S1. Demonstration of peptide to protein summarization for two candidate genes.**

363 A) Histogram shows the DUX4/Control ratios (log2) for all peptides

364 B-C) Histogram shows the DUX4/control ratios (log2) for all peptides mapping to RPL15 (B) or

365 ZSCAN4 gene (C)

366 D) Box and whisker plot showing the median log2 DUX4/control ratio for RPL15 and ZSCAN4.

367

368 **Figure S2. Tool for easy access of the data generated in this study**

369 Screenshot of the Shiny web server showing a sample analysis of the RNA and protein-level data

370 for a DUX4 transcriptional target, ZSCAN4.

371

372

373 **Supplementary Data**

374 Supplementary data 1: PSM-level data (deposited to Dryad; Link:

375 https://datadryad.org/review?doi=doi:10.5061/dryad.ck06k75)

376 Supplementary data 2: Gene level data for RNA and protein fold-change

377 Supplementary file 3: R markdown document with the code needed to process mass spectrometry

378 data and generate figures.

379

380

Acknowledgements

The authors wish to acknowledge Heidi Dvinge for helpful discussions on proteomics data analysis. We also thank current and past members of the Bradley and Tapscott laboratories for comments and suggestions that have improved this body of work over time. This research was supported by the NINDS P01NS069539 (RKB and SJT), and the FSH Society FSHS-22014-01 (SJ). RKB is a Scholar of The Leukemia & Lymphoma Society. The Fred Hutchinson Cancer Research Center Proteomics Facility is supported by a National Institutes of Health Cancer Center Support Grant (P30 CA015704). The OrbiTrap Fusion mass spectrometer used in this research was purchased with a grant from the M.J. Murdock Charitable Trust.

Competing interests

No competing interests declared.

References

1. Tawil R, van der Maarel SM, & Tapscott SJ (2014) Facioscapulohumeral dystrophy: the path to consensus on pathophysiology. *Skelet Muscle* 4:12.
2. Lemmers RJ*, et al.* (2010) A unifying genetic model for facioscapulohumeral muscular dystrophy. *Science* 329(5999):1650-1653.
3. Geng LN*, et al.* (2012) DUX4 activates germline genes, retroelements, and immune mediators: implications for facioscapulohumeral dystrophy. *Dev Cell* 22(1):38-51.
4. Block GJ*, et al.* (2013) Wnt/beta-catenin signaling suppresses DUX4 expression and prevents apoptosis of FSHD muscle cells. *Hum Mol Genet* 22(23):4661-4672.
5. Young JM*, et al.* (2013) DUX4 binding to retroelements creates promoters that are active in FSHD muscle and testis. *PLoS Genet* 9(11):e1003947.
6. Banerji CR*, et al.* (2015) beta-Catenin is central to DUX4-driven network rewiring in facioscapulohumeral muscular dystrophy. *J R Soc Interface* 12(102):20140797.
7. Feng Q*, et al.* (2015) A feedback loop between nonsense-mediated decay and the retrogene DUX4 in facioscapulohumeral muscular dystrophy. *Elife* 4.
8. Homma S, Beermann ML, Boyce FM, & Miller JB (2015) Expression of FSHD-related DUX4-FL alters proteostasis and induces TDP-43 aggregation. *Ann Clin Transl Neurol* 2(2):151-166.
9. Dmitriev P*, et al.* (2016) DUX4-induced constitutive DNA damage and oxidative stress contribute to aberrant differentiation of myoblasts from FSHD patients. *Free Radic Biol Med* 99:244-258.
10. Shadle SC*, et al.* (2017) DUX4-induced dsRNA and MYC mRNA stabilization activate apoptotic pathways in human cell models of facioscapulohumeral dystrophy. *PLoS Genet* 13(3):e1006658.
11. Campbell AE, Belleville A, Resnick R, Shadle SC, & Tapscott SJ (2018) Facioscapulohumeral dystrophy: Activating an early embryonic transcriptional program in human skeletal muscle. *Hum Mol Genet*.
12. Lek A, Rahimov F, Jones PL, & Kunkel LM (2015) Emerging preclinical animal models for FSHD. *Trends Mol Med* 21(5):295-306.
13. Tassin A*, et al.* (2013) DUX4 expression in FSHD muscle cells: how could such a rare protein cause a myopathy? *J Cell Mol Med* 17(1):76-89.
14. Rickard AM, Petek LM, & Miller DG (2015) Endogenous DUX4 expression in FSHD myotubes is sufficient to cause cell death and disrupts RNA splicing and cell migration pathways. *Hum Mol Genet* 24(20):5901-5914.
15. Knopp P*, et al.* (2016) DUX4 induces a transcriptome more characteristic of a less-differentiated cell state and inhibits myogenesis. *J Cell Sci* 129(20):3816-3831.
16. Schwanhausser B*, et al.* (2011) Global quantification of mammalian gene expression control. *Nature* 473(7347):337-342.
17. Tassin A*, et al.* (2012) FSHD myotubes with different phenotypes exhibit distinct proteomes. *PLoS One* 7(12):e51865.
18. Celegato B*, et al.* (2006) Parallel protein and transcript profiles of FSHD patient muscles correlate to the D4Z4 arrangement and reveal a common impairment of slow to fast fibre differentiation and a general deregulation of MyoD-dependent genes. *Proteomics* 6(19):5303-5321.
19. Laoudj-Chenivesse D*, et al.* (2005) Increased levels of adenine nucleotide translocator 1 protein and response to oxidative stress are early events in facioscapulohumeral muscular dystrophy muscle. *J Mol Med (Berl)* 83(3):216-224.
20. Harsha HC, Molina H, & Pandey A (2008) Quantitative proteomics using stable isotope labeling with amino acids in cell culture. *Nat Protoc* 3(3):505-516.

21. Jagannathan S*, et al.* (2016) Model systems of DUX4 expression recapitulate the transcriptional profile of FSHD cells. *Hum Mol Genet* 25(20):4419-4431.
22. Yao Z*, et al.* (2014) DUX4-induced gene expression is the major molecular signature in FSHD skeletal muscle. *Hum Mol Genet* 23(20):5342-5352.
23. Rodriguez-Boulan E & Musch A (2005) Protein sorting in the Golgi complex: shifting paradigms. *Biochim Biophys Acta* 1744(3):455-464.
24. Bexiga MG & Simpson JC (2013) Human diseases associated with form and function of the Golgi complex. *Int J Mol Sci* 14(9):18670-18681.
25. Claudio N, Dalet A, Gatti E, & Pierre P (2013) Mapping the crossroads of immune activation and cellular stress response pathways. *EMBO J* 32(9):1214-1224.
26. Oslowski CM & Urano F (2011) Measuring ER stress and the unfolded protein response using mammalian tissue culture system. *Methods Enzymol* 490:71-92.
27. Palacios IM (2013) Nonsense-mediated mRNA decay: from mechanistic insights to impacts on human health. *Brief Funct Genomics* 12(1):25-36.
28. Licklider LJ, Thoreen CC, Peng J, & Gygi SP (2002) Automation of nanoscale microcapillary liquid chromatography-tandem mass spectrometry with a vented column. *Anal Chem* 74(13):3076-3083.
29. Kall L, Canterbury JD, Weston J, Noble WS, & MacCoss MJ (2007) Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat Methods* 4(11):923-925.

Figure 1

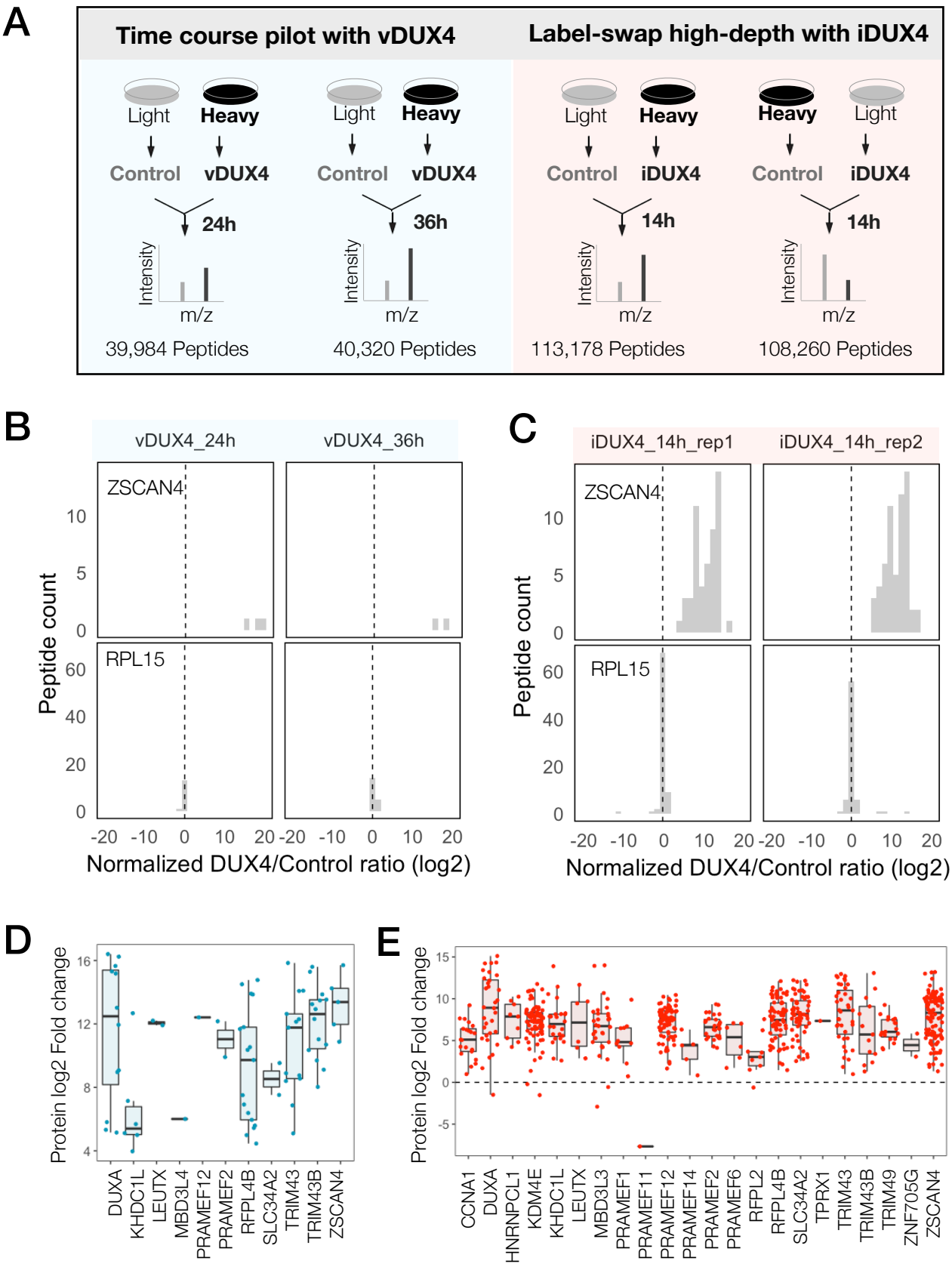
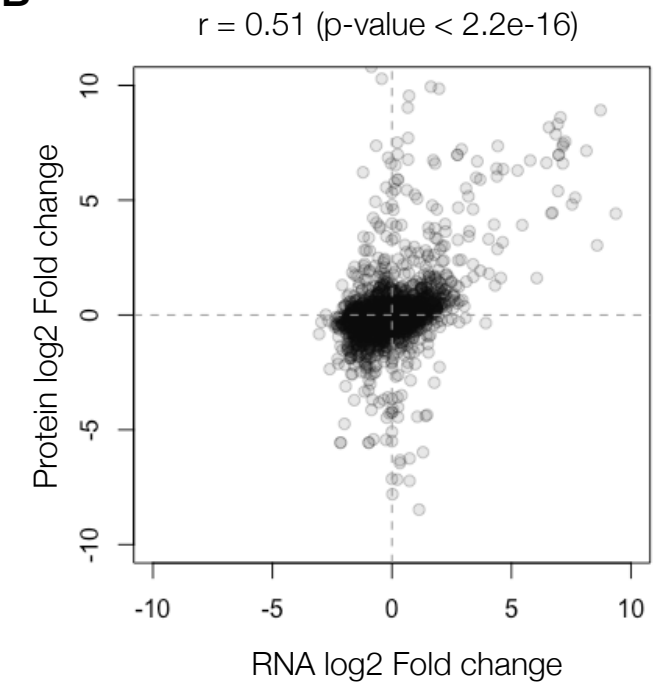


Figure 2

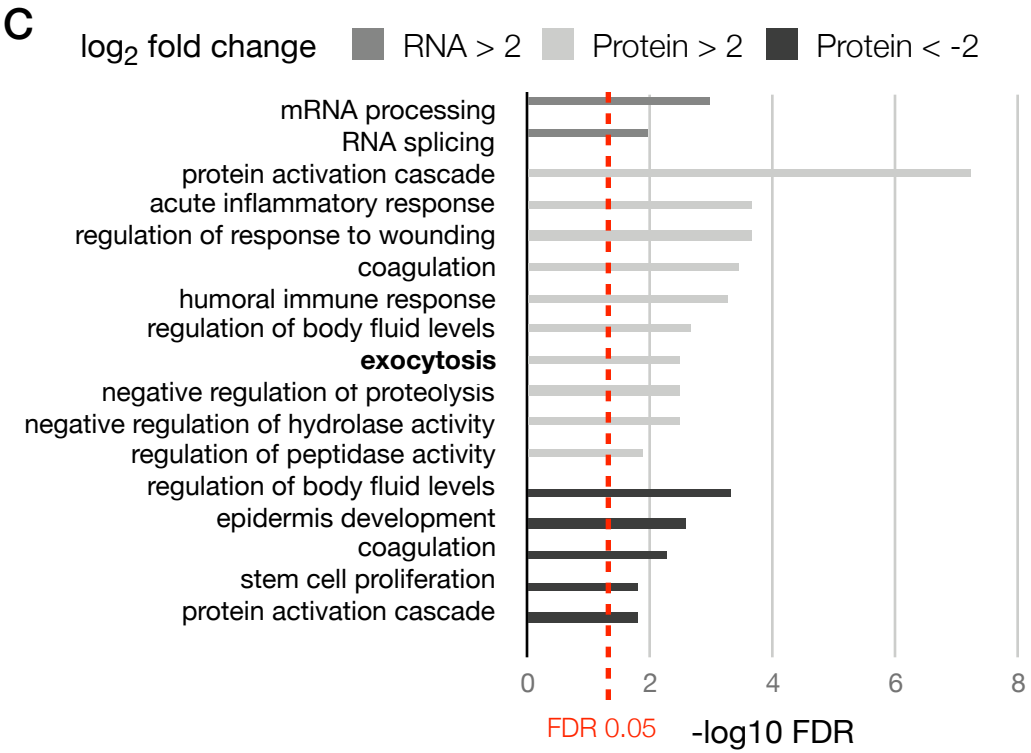
A

	Gene number		
Criteria	**RNA**	**Protein**	**Both**
Measured	21184	4005	3961
log2 fold change > 2	107	132	54
log2 fold change < -2	36	76	3

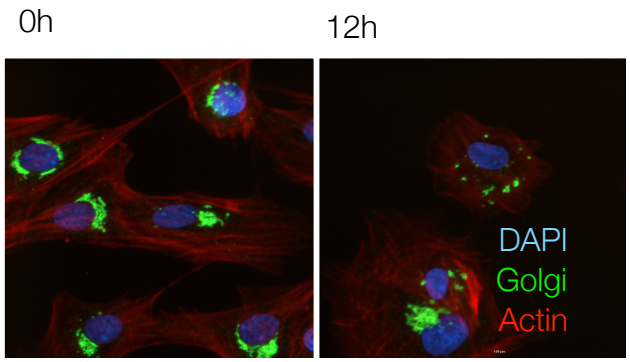
B



C



D



E

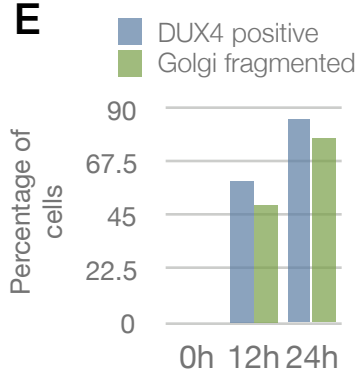


Figure 3

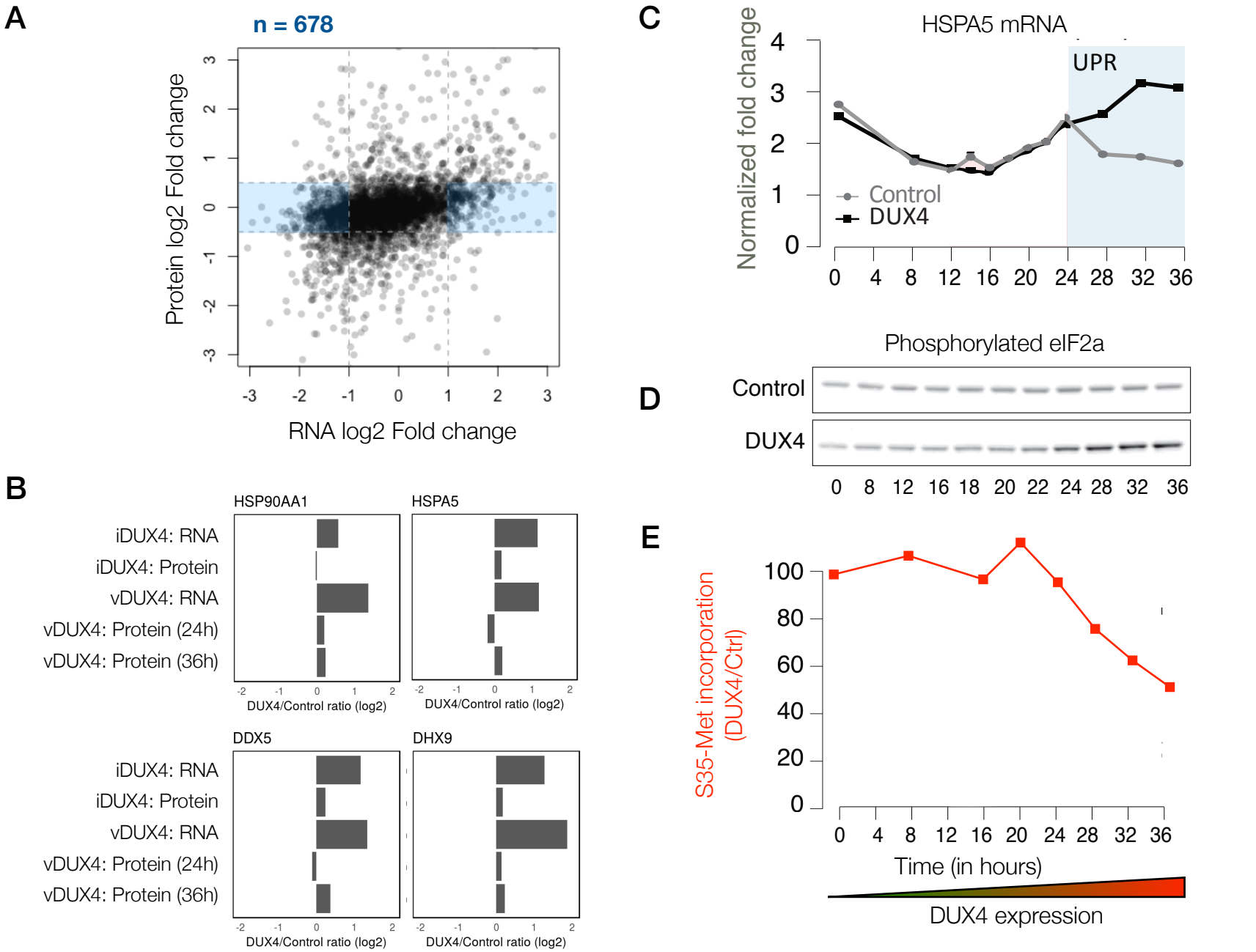

A
n = 678
Protein log2 Fold change
RNA log2 Fold change
B
HSP90AA1
HSPA5
iDUX4: RNA
iDUX4: Protein
vDUX4: RNA
vDUX4: Protein (24h)
vDUX4: Protein (36h)
DUX4/Control ratio (log2)
DDX5
DHX9
C
HSPA5 mRNA
UPR
Normalized fold change
Control
DUX4
D
Phosphorylated eIF2a
Control
DUX4
E
S35-Met incorporation (DUX4/Ctrl)
Time (in hours)
DUX4 expression

Figure 4

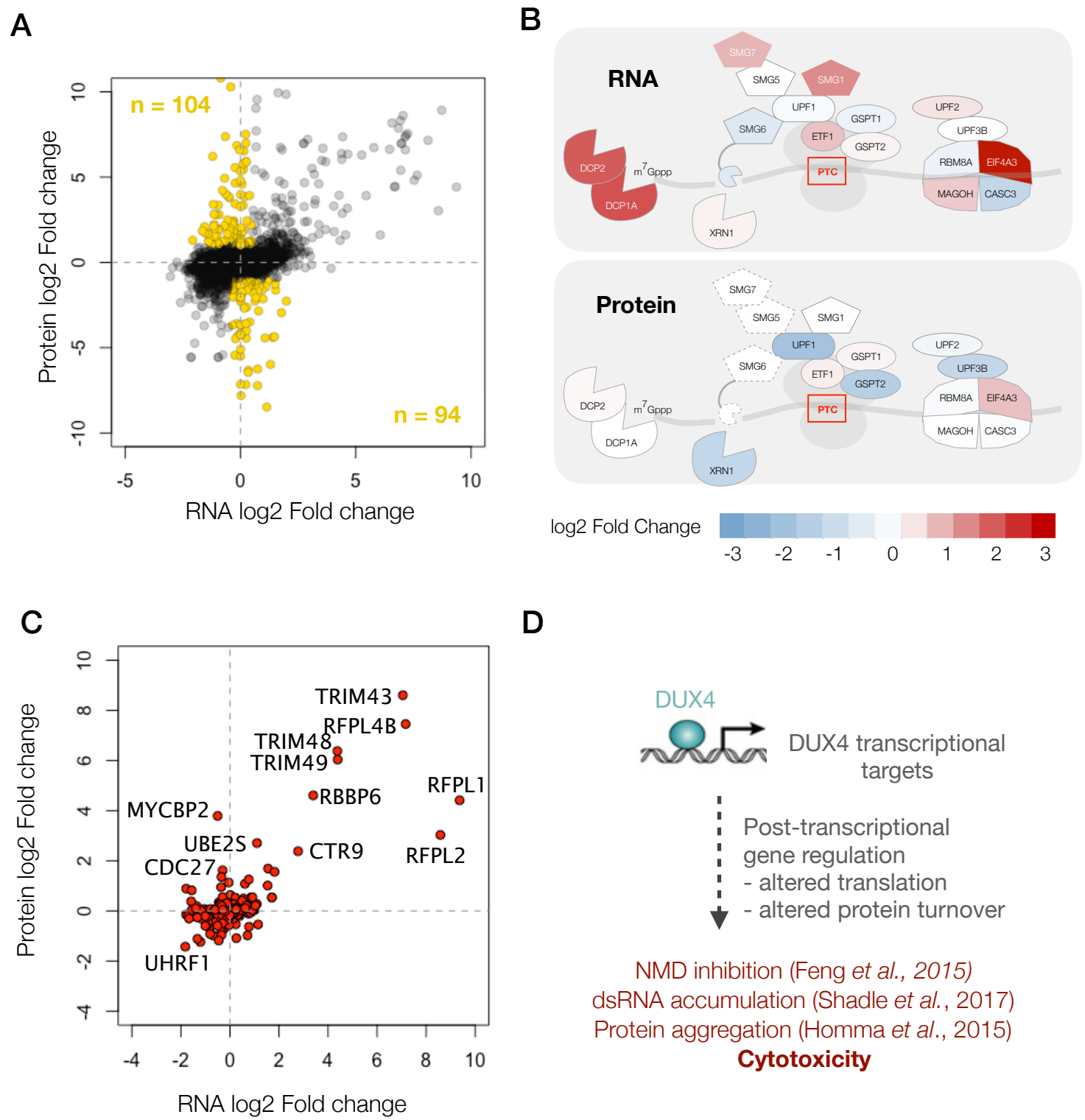


Figure S1

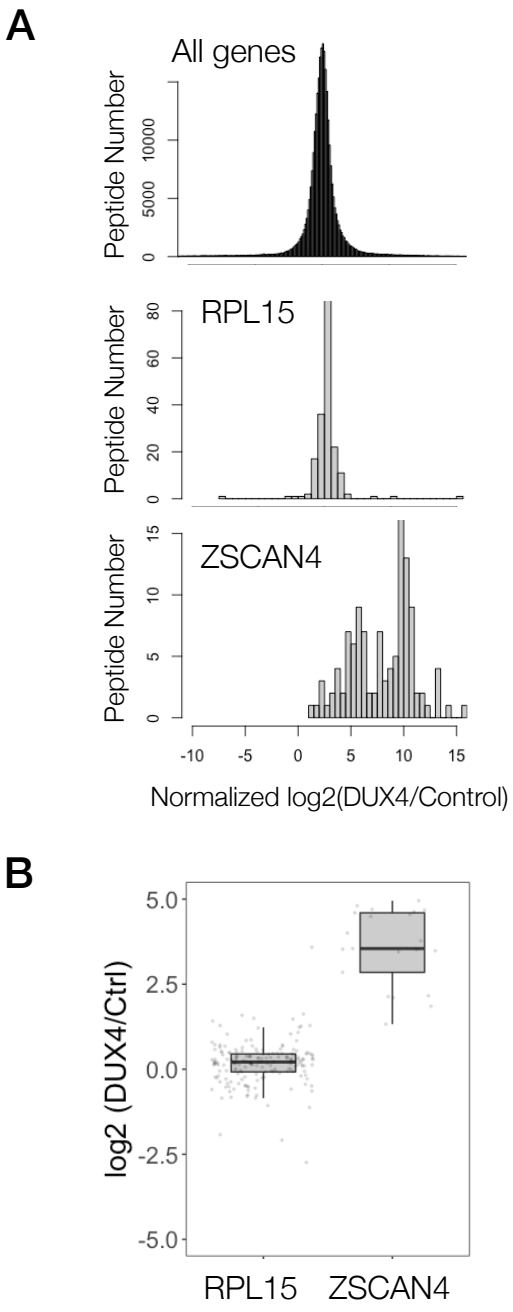
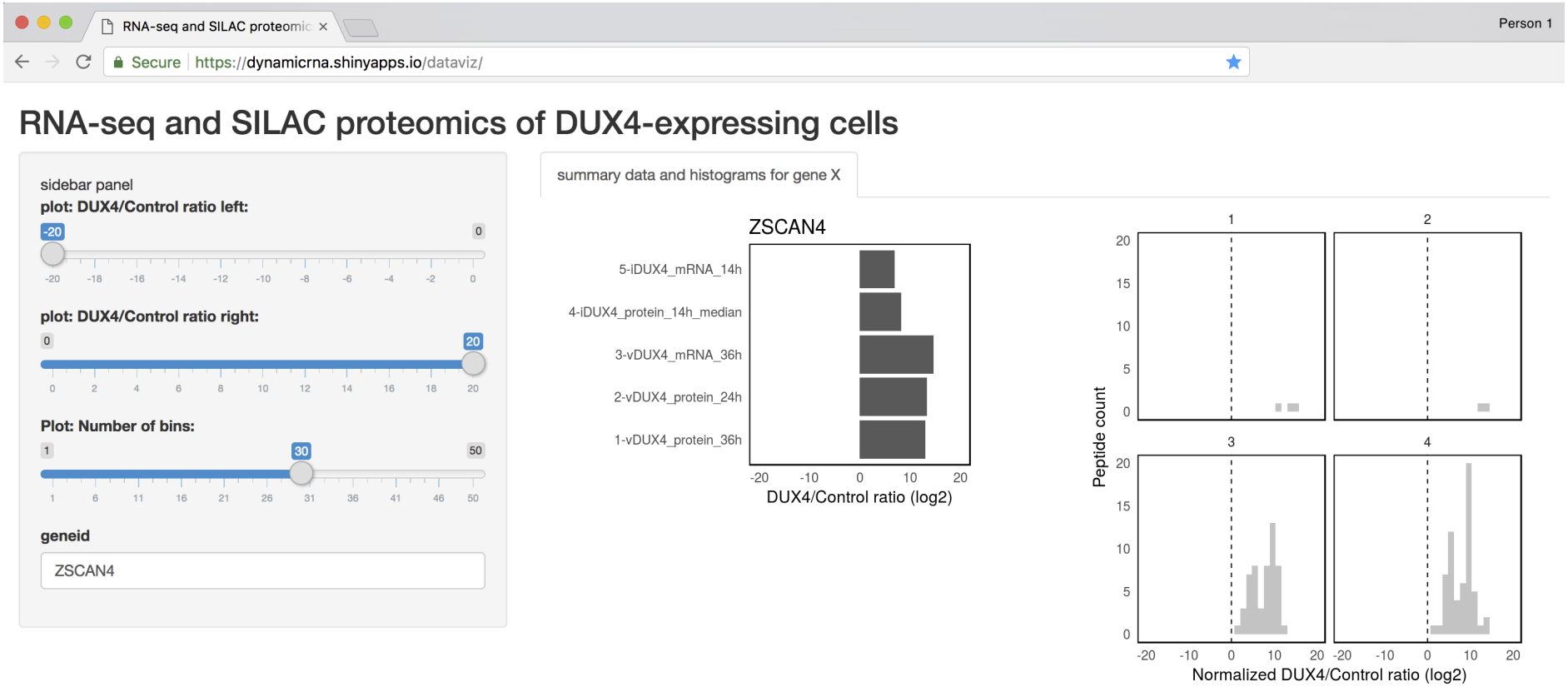


Figure S2


RNA-seq and SILAC proteomic
Person 1
Secure https://dynamicrna.shinyapps.io/dataviz/
RNA-seq and SILAC proteomics of DUX4-expressing cells
sidebar panel
plot: DUX4/Control ratio left:
-20
0
plot: DUX4/Control ratio right:
0
20
Plot: Number of bins:
1
30
50
geneid
ZSCAN4
summary data and histograms for gene X
ZSCAN4
5-iDUX4_mRNA_14h
4-iDUX4_protein_14h_median
3-vDUX4_mRNA_36h
2-vDUX4_protein_24h
1-vDUX4_protein_36h
DUX4/Control ratio (log2)
Peptide count
Normalized DUX4/Control ratio (log2)