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

[2]Basic 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 Shared Resource, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA

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

*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 machineries, 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 processes to DUX4-induced pathology.

38 **Introduction**

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

47 DUX4 induces changes in the expression of hundreds of genes that impact dozens of
48 highly interconnected pathways (3-10), making a cause-and-effect relationship between
49 dysregulated gene expression and FSHD pathology difficult to discern. Because DUX4 is a strong
50 transcriptional activator, most studies of DUX4 activity have focused on measuring gene
51 expression at the transcript level (3, 14, 15), thereby implicitly assuming that the transcriptome
52 accurately represents the cellular proteome in DUX4-expressing cells. While this is a reasonable
53 assumption, it is well known that RNA and protein levels are not always concordant and that post-
54 transcriptional regulation can results in divergent RNA and protein levels (16). While a few
55 proteomics studies have been conducted on FSHD muscle biopsies, these early studies lack the
56 depth necessary to make meaningful comparisons with the DUX4-induced transcriptome (17-19).
57 Furthermore, given our recent discovery that DUX4 induces proteolysis of a key RNA binding
58 protein, UPF1 (7), we hypothesized that conducting paired measurements of RNA and protein
59 levels might be particularly important in order to identify altered post-transcriptional gene
60 regulation in DUX4-expressing cells. Hence, we set out to generate reliable RNA- and protein-
61 level measurements of DUX4-induced gene expression and thereby elucidate the extent of post-
62 transcriptional gene dysregulation in DUX4-expressing cells.

63 We used our previously established and validated cell culture models of DUX4 expression
64 (20) to conduct RNA-seq and Stable Isotope Labeling with Amino acids in Cell culture (SILAC)
65 coupled with quantitative mass spectrometry (21). The resulting data enabled us to measure
66 DUX4-induced alterations in the transcriptome as well as protein levels for ~4,000 genes with
67 high confidence. Comparing the transcript-level and protein-level alterations revealed three
68 distinct patterns of expression for different subsets of genes: 1) Concordant changes in
69 expression at the RNA and protein level for many transcriptional targets of DUX4; 2) Post-
70 transcriptional buffering of the expression of many genes, especially of those involved in stress
71 response pathways; and 3) Discordant gene expression changes at the RNA versus protein levels
72 for many genes, particularly those involved in RNA surveillance. Together, these findings highlight
73 the importance of measuring the expressed proteome in order to fully understand DUX4 biology
74 and the FSHD disease process.

75 **Results**

76 **Determining protein-level alterations in DUX4-expressing cells via quantitative mass**
77 **spectrometry.**

78 In order to measure DUX4-induced changes to the cellular proteome, we conducted
79 SILAC-based mass spectrometry in two independent DUX4 expression systems (**Figure 1A).** We
80 previously showed that DUX4 expressed via a lentiviral vector versus an inducible transgene
81 integrated into the genome of a myoblast cell line both yield comparable gene expression profiles
82 that accurately capture the transcriptome of FSHD cells (20). Here, we used both of these
83 expression systems in order to internally corroborate our results, thereby ensuring that our
84 proteomic data were robust with respect to choice of model system.

85 We first established the efficiency of SILAC labeling and determined appropriate cell
86 culture conditions and experimental timing in order to ensure near-complete labeling of proteins
87 with heavy isotopic Arginine and Lysine. To this end, we cultured immortalized MB135 myoblasts
88 in SILAC labeling media for a week and subjected the total protein from labeled cells, as well as
89 a 1:1 mix of labeled to unlabeled cells, to mass spectrometry. Using the resulting mass spectra,
90 we calculated labeling efficiency in two ways. First, we quantified the relative abundance of heavy
91 and light spectra for several individual peptides and found that it showed a >95% labeling
92 efficiency (a representative example shown in **Figure S1A**). Next, we analyzed the proteome-
93 wide distribution of log-transformed heavy to light ratios of peptides from both the heavy-only
94 (**Figure S1B**) and heavy:light mixed samples (**Figure S1C**). The distribution of heavy-to-light
95 ratios was strongly skewed to the right in the heavy-only samples, indicating robust labeling with
96 the heavy amino acids. In contrast, the heavy:light mixed sample yielded a distribution centered
97 roughly around zero, as expected. Based on these data, we proceeded with the labeling
98 conditions and increased the labeling time even further (to three weeks) in order to achieve
99 maximal SILAC labeling of the MB135 myoblasts.

100 MB135 myoblasts were adapted to light or heavy SILAC media for 3 weeks were
101 transduced with lentivirus carrying DUX4 (vDUX4) or GFP (vGFP) expression constructs.
102 Samples were collected 24- and 36-hours post-transduction. In an independent experiment,
103 MB135 cells carrying a doxycycline-inducible DUX4 transgene (iDUX4; (20)) were adapted to
104 SILAC media for 3 weeks and DUX4 expression was induced with 1μg/ml of doxycycline for 14
105 hours in two replicates carrying heavy and light SILAC labels. Paired controls with no treatment
106 were also collected with both heavy and light labels. Total protein from cells expressing DUX4
107 were mixed with an equal amount of total protein from cells without DUX4 expression containing
108 the opposite SILAC label to generate samples that were then subjected to mass spectrometry.

109 Peptide-spectrum matches (PSMs) with quantified heavy to light ratios were subject to
110 thorough screening for quality (e.g., filtering out single-peak spectra and spectra without unique
111 mapping; **Figure 1A**; see *Materials & Methods* for further details; Supplemental Data Set 1). A
112 histogram of log-transformed DUX4 to Control abundance ratios of peptides mapping to a DUX4
113 target gene, ZSCAN4, from both vDUX4 and iDUX4 datasets showed highly skewed $\log_2$
114 (DUX4/Control) ratio, consistent with significant upregulation of the protein upon DUX4
115 expression (**Figure 1B-C**). In contrast, plotting the $\log_2$ (DUX4/Control) ratio of all individual
116 peptides mapping to the housekeeping gene RPL15 showed that the ratio is centered around
117 zero (**Figure 1B-C**), as would be expected for a gene with no differential expression upon DUX4
118 induction. These example plots illustrate the strong agreement between the expected and
119 observed protein fold change values determined by SILAC mass spectrometry. Moreover, out of
120 the 65 genes identified by Yao *et al* (22) as potential FSHD biomarkers based on transcriptome
121 analysis of FSHD patient samples, 8 were quantified in the vDUX4 proteomics study and 25 were
122 quantified in the iDUX4 proteomics study and both show high induction at the protein level (**Figure
123 1D, E**). Note that the lower number of peptides (and hence quantified proteins) from the vDUX4

124 sample indicates the lower depth of this dataset, and yet yields fold changes that are highly
125 consistent with the iDUX4 dataset with a higher depth.

126

127 **Assessing concordance of RNA and protein expression fold change.**

128 Next, using the iDUX4 dataset, we performed peptide to protein summarization via median
129 heavy/light ratios of all the peptides mapping to a certain protein in both replicates to obtain gene-
130 level $\log_2$ (DUX4/Control) ratios (**Figure 1A and S1D-E**; Supplemental Data Set 2). After filtering
131 out genes that were only observed in one of the two label-swap replicates, we obtained
132 quantitative proteomics information for 4005 genes, 3961 of which also had a corresponding RNA-
133 seq measurement (**Figure 2A;** RNA-seq data previously reported in (20)). The lower number of
134 genes quantified via proteomics compared to RNA-seq is expected as proteomics is known to
135 have lower sensitivity compared to RNA-seq.

136 To qualitatively compare the RNA and protein expression level changes upon DUX4
137 expression, we assessed the overlap of genes with an expression change of 4-fold or above.
138 Among genes that are upregulated (> 2 log2 fold change), the concordance between RNA and
139 protein was roughly 40-50%, whereas similarly downregulated genes show very little concordance
140 (**Figure 2A**). To obtain a more quantitative measure of concordance, we generated a scatter plot
141 of the RNA versus protein fold change for the 3961 genes (**Figure 2B**). We found a reasonable
142 level of correlation between these values with a Pearson's correlation coefficient, r, of 0.51 (p-
143 value < 2.2e-16).

144 To assess whether similar pathways were affected at the RNA versus protein level, we
145 conducted Gene Ontology (GO) analysis for genes up- or down-regulated at the RNA and protein
146 (**Figure 2C**). Surprisingly, we observed that the pathways that are affected at the RNA versus
147 protein levels are quite distinct. While the transcript level changes occur in genes involved in
148 transcription and mRNA processing, protein-level changes impact pathways including humoral
149 immune response, proteolysis and exocytosis. As the exocytosis pathway has not been implicated

150 in any of the previous DUX4 gene expression studies, we sought to examine this phenomenon
151 further by imaging the Golgi apparatus, which is the source of exocytotic vesicles in the cell (23).
152 We found that DUX4 expressing cells showed severe fragmentation of Golgi apparatus, which
153 could be an indicator of a perturbation in the cellular secretory pathways (24) (**Figure 2D;**
154 quantified in **Figure 2E**).

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

158

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

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

166 Given that both unfolded protein and dsRNA induced stresses converge in
167 phosphorylation of eIF2α and lead to translation inhibition (25), we asked if the timing of
168 transcription of various stress-response genes coincide with translational downregulation. We
169 found that HSPA5, a prominent marker of the unfolded protein response pathway (26), shows
170 transcriptional upregulation at a time period that temporally coincides with eIF2α phosphorylation
171 and reduced incorporation of [^{35}S]-labeled methionine, a proxy for bulk translation efficiency
172 (**Figure 3 C-E**). These data demonstrate that translation inhibition caused by various cellular
173 stresses and the resulting post-transcriptional buffering prevents DUX4-expressing cells from
174 mounting a robust stress response.

175

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

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

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

192 Post-transcriptional downregulation of a gene can be achieved via two means: reduced
193 translation or increased protein degradation. We have previously shown that DUX4 induces
194 proteasome-mediated degradation of UPF1 (7). Hence, we asked if DUX4 affects known
195 components and regulators of the ubiquitin proteasome. A scatterplot of all ubiquitin proteasome
196 regulators shows a change in the expression of several such genes, one or more of which may
197 underlie the rapid degradation of UPF1 (**Figure 4C**). Further studies are needed to understand
198 the precise molecular mechanism behind this regulatory pathway and its downstream

199 consequences. In summary, we propose that the post-transcriptional gene regulation plays a
200 critical role in inhibiting NMD and perturbing the proteostasis in DUX4-expressing cells and may
201 thus underlie key aspects of FSHD pathology (**Figure 4D**).

202 Finally, in order to enable researchers and patients in the FSHD community to access the
203 data generated in this study, we developed a web tool for easy visualization of these data
204 (Screenshot shown in **Figure S2**). This tool can be freely accessed at the following URL:
205 https://dynamicrna.shinyapps.io/dataviz/.
206
207
208

209 **Discussion**

210 Most of the highly induced DUX4 transcriptional targets are germline and early embryonic
211 genes that are normally never expressed in adult muscle (3). So, it is possible that despite being
212 expressed at the transcript level, such genes may be translated poorly and/or be degraded rapidly
213 upon translation due to the lack of cell-type chaperones or other factors. Here, we used
214 quantitative mass spectrometry in two different cell culture models of FSHD to demonstrate that
215 DUX4-induced transcripts are efficiently translated into stable proteins in muscle cells. As we
216 previously demonstrated that our vDUX4 and iDUX4 systems accurately capture the
217 transcriptional program of FSHD cells (20), it is reasonable to assume that DUX4-induced
218 transcripts are similarly translated into stable proteins in FSHD muscle. However, further work is
219 required to definitively confirm this. Similarly, our data strongly motivate focused investigation of
220 these DUX4-induced proteins to test their utility as potential FSHD biomarkers in relevant clinical
221 samples (22).

222 Next, we asked if the changes in the DUX4 proteome are largely reflective of the changes
223 to the transcriptome. We found that this is not the case. While there is a positive correlation
224 between these measurements (Pearson's correlation coefficient of 0.51), hundreds of genes
225 deviate from this trend. GO analysis of the most differentially expressed genes at the transcript
226 versus protein levels revealed RNA splicing and processing as being the prominent categories
227 impacted at the transcript level, while protein-level changes impacted an entirely different set of
228 pathways. We take these results as indication that the transcriptome-level analysis paints an
229 incomplete picture of DUX4 biology that should be complemented with proteome-level analysis
230 to thoroughly understand how DUX4 misexpression causes FSHD.

231 We next pursued the various mechanisms by which protein-level changes deviate from
232 the corresponding transcriptomic changes. We found that genes induced by dsRNA and unfolded
233 protein stress are transcriptionally induced, but translationally buffered as a result of the

234 translational repression that accompanies activation of these stress-response pathways.
235 Motivated by this result, we hypothesize that DUX4-expressing cells are unable to mount a robust
236 stress response despite inducing the transcripts necessary to alleviate stress. We also show that
237 multiple proteins in RNA surveillance pathways, including UPF1 and XRN1, are downregulated
238 at the protein level, which may explain the drastic reduction in RNA quality control capacity that
239 we observed in DUX4-expressing cells. From the proteomics data, we identified many genes
240 involved in the ubiquitin proteasome pathway whose mis-regulation could alter protein stability.
241 These genes may serve as a starting point for further investigation of how DUX4, which is best
242 known as a transcriptional activator, induces widespread alterations in protein turnover.

243 **Materials and Methods**

244 **Accession codes.** Mass spectrometry proteomics data have been deposited to the
245 ProteomeXchange Consortium via the PRIDE partner repository (Vizcaino et al., 2013) with the
246 dataset identifier PXD010221 (Reviewer account details: Username: reviewer37110@ebi.ac.uk;
247 Password: aMrE4BW0). The RNA-seq data are available through the NCBI SRA database under
248 accession number GSE85461 (20).

249

250 **Cell culture and SILAC labeling.** Proliferating human myoblasts were cultured in F10 medium
251 (Gibco/Life Technologies) supplemented with 20% fetal bovine serum (Thermo Scientific), 10ng
252 bFGF (Life Technologies), 1μM dexamethasone (Sigma) and 50U/50μg penicillin/streptomycin
253 (Life Technologies). Cells were labeled in SILAC media containing heavy lysine (Lys8) and
254 arginine (Arg10) for 3 weeks before DUX4 induction experiments. To induce DUX4 expression in
255 the MB135 iDUX4 cells, 1μg/ml of doxycycline was added for 8 or 14 hours, as indicated. For viral
256 DUX4 expression, MB135 cells were transduced with lentivirus carrying DUX4 coding sequence
257 under a hPGK promoter in the presence of polybrene.

258

259 **Gel slice digestion.** Total RNA and protein were extracted from whole cells using TRIzol^{RT}
260 reagent (Ambion) following the manufacturer's instructions. 50 μg of total protein was subjected
261 to SDS PAGE using a 4-15% bis-TRIS gel. The gel was stained using GelCode blue (Pierce) as
262 per manufacturer's instructions, destained overnight in ultrapure water and the entire lane
263 containing protein was cut into 16 fractions using a GelCutter (Gel company Inc.). Individual gel
264 slices in 1.5 mL tubes (Eppendorf) were consecutively washed with water and incubated with
265 25mM ammonium bicarbonate, 50% acetonitrile for 2 hrs. The gel pieces were dehydrated with
266 acetonitrile. And the dried gel slices were reduced by covering them with 10 mM dithiothreitol in
267 100 mM ammonium bicarbonate and heating them at 56°C for 45 min. The solution was removed

268 and discarded. The gel slices were alkylated by covering them with a solution of 50 mM
269 iodoacetamide in 100 mM ammonium bicarbonate and incubating in the dark at ambient
270 temperature for 30 min. The solution was removed and discarded. The gel slices were dehydrated
271 with acetonitrile, then washed with 100 mM ammonium bicarbonate 100 mM for 10 min. The
272 solution was removed, discarded and the gel slices were dehydrated once again with acetonitrile.
273 After removing acetonitrile, the gel slices were then hydrated with 5 ng/uL sequencing grade
274 trypsin (Promega) in 50 mM ammonium bicarbonate and digested overnight at 37°C on an orbital
275 shaker. Following digestion, the supernatants were collected, and the gel slices were washed with
276 0.1% trifluoroacetic acid, and after 30 min an equal volume of acetonitrile was added followed by
277 washing for an additional 1 hour. The original digestion supernatant and the wash for a single
278 sample were combined into a single tube and dried by vacuum centrifugation. The digestion
279 products were desalted using Ziptips (Millipore) per the manufacturer's instructions, eluted with
280 70% acetonitrile/0.1% trifluoroacetic acid, and dried by vacuum centrifugation.

281

282 **Mass spectrometry.** The desalted material was resuspended in 20 µL of 2% acetonitrile in 0.1%
283 formic acid, and 18 µL was analyzed using one of two LC/ESI MS/MS configurations. The first
284 configuration consisted of an Easy-nLC II (Thermo Scientific) coupled to a Orbitrap Elite ETD
285 (Thermo Scientific) mass spectrometer using a trap-column configuration as described (28). A
286 trap of 100 µm × 20 mm packed with Magic C_{18}AQ (5-µm, 200 Å resin; Michrom Bioresources)
287 packing material was used for in-line desalting and a column of 75 µm × 250 mm packed with
288 C_{18}AQ (5-µm, 100 Å resin; Michrom Bioresources) was used for analytical peptide separations.
289 Chromatographic separations were carried out using a 60-minute gradient from 5% to 35%
290 solvent B (solvent A: 0.1% formic acid, solvent B: 0.1% formic acid in acetonitrile) at a flowrate of
291 300 nL/min. The analytical column temperature was maintained at 40℃. The Orbitrap Elite
292 instrument was operated in the data-dependent mode, switching automatically between MS

293 survey scans in the Orbitrap (AGC target value 1E6, resolution 240,000, and maximum injection
294 time 250 ms) and collision induced dissociation (CID) MS/MS spectra acquisition in the linear ion
295 trap (AGC target value of 10,000 and injection time 100 ms). The twenty most intense precursor
296 ions from the OrbiTrap full scan were each consecutively selected for fragmentation by CID in the
297 linear ion trap using a normalized collision energy of 35%. Ions of +2 and +3 charge states were
298 selected for MS/MS and selected ions were dynamically excluded for 30 seconds. The second
299 configuration consisted of an Easy nanoLC 1000 (Thermo Scientific) HPLC connected to an
300 OrbiTrap Fusion (Thermo Scientific) mass spectrometer. In-line chromatographic separations (no
301 trap column) were carried out using a 75 µm × 400 mm column packed with Magic C_{18}AQ (5-µm,
302 100 Å resin; Michrom Bioresources) packing material at a flowrate of 300 nL/min.
303 Chromatographic elution consisted of a 90-minute gradient from 3% to 27% B and the column
304 temperature was maintained at 40℃. The OrbiTrap Fusion was operated in the 2 second "top
305 speed" data dependent acquisition mode with MS survey scans in the OrbiTrap at least every 2
306 seconds (AGC target value 4E5, resolution 120,000, and maximum injection time of 50 ms).
307 Quadrupole isolation was set to 1.6 FWHM and higher energy collision dissociation (HCD) was
308 used for fragmentation at a collision energy of 28% and MS/MS detection was carried out in the
309 linear ion trap set at rapid scan speed (injection time of 250 ms and AGC target of 10E2).
310 Positively charged ions from 2 to 6 were selected for MS/MS and selected ions were dynamically
311 excluded for 30 seconds.

312

313 **Data analysis and statistical methods.** Qualitative and quantitative data analysis were
314 performed using Proteome Discoverer 2.1 (Thermo Scientific). The data were searched against
315 a human UniProt database (downloaded 11-04-16) that was appended with protein sequences
316 from the common Repository of Adventitious Proteins (cRAP; www.thegpm.org/crap/) an *in silico*
317 translation products of noncanonical transcript isoforms stabilized due to NMD inhibition. In

318 downstream analyses, peptides that only mapped to NMD targets were not considered any further
319 in the current study and will be pursued in a future investigation. Searches were conducted with
320 the trypsin enzyme specificity. The precursor ion tolerance was set to 10 ppm and the fragment
321 ion tolerance was set to 0.6 Da. Variable modifications were set for oxidation on methionine
322 (+15.995 Da), carbamidomethyl (+57.021 Da) on cysteine, and acetylation (+42.010 Da) on the
323 N-terminus of proteins. Heavy SILAC amino acids for lysine (+8.014 Da) and arginine (10.008)
324 were also accounted for in the analysis as variable modifications. All search results were
325 evaluated by Percolator (29) for false discovery rate (FDR) evaluation of the identified peptides.
326 Peptide identifications were filtered to a peptide FDR of 1%.

327 The peptide-spectrum matches (PSMs) and the corresponding quantification data were
328 then exported as a tab-delimited text file from Proteome Discoverer 2.1, and all downstream data
329 analysis was conducted using the R statistical programming language. The Custom R script
330 (**Supplementary File 3**) to reproduce the analyses described above and figures in this manuscript
331 is deposited in github, available at the following URL: https://github.com/sjaganna/2018-
332 jagannathan_et_al. Briefly, we filtered PSMs to remove the spectra carrying the following quality
333 flags assigned by Proteome Discoverer – "Inconsistently labeled", "Indistinguishable channels",
334 "No quantitative values", "Not unique", "Redundant spectra", "Single peak spectra", and "Excluded
335 by method" – all of which represent poor quality spectra. Next, we removed peptides that failed
336 to map to a gene, and peptides mapping to genes in the cRAP database or to uncharacterized
337 genes. We then assigned the median peptide fold change within an experiment as the
338 summarized protein fold change value.

339 Upon comparing the gene-level fold changes obtained from four different samples, we
340 noticed a strong correlation between the two viral replicates (Pearson's r = 0.81), as well as the
341 data from iDUX4 replicate 1 and both vDUX4 datasets (Pearson's r = 0.79 and 0.77). The iDUX4
342 replicate 2, in which we swapped the isotopic labels used for the control and DUX4 samples,

343 however, showed poor correlation with the three other samples due to a small set of proteins (103
344 proteins; 2.6% of total) that did not label efficiently. Proteins that do not label efficiently result in
345 apparent anticorrelation between label-swap samples. This issue, wherein a small subset of
346 proteins does not label efficiently, is a well-known problem in SILAC experiments (30). Rather
347 than ignore this subset of proteins, which would require the imposition of potentially arbitrary
348 thresholds, we instead followed the standard SILAC analysis procedure of taking the median
349 value of H/L ratio across the two iDUX4 replicates as the final iDUX4 H/L ratio (30, 31). This
350 procedure preserves the H/L ratio for proteins that labeled efficiently (the vast majority of the
351 proteome) while bringing the H/L ratio for proteins that did not label efficiently to approximately 1
352 (corresponding to no measurable change). We then calculated *p*-values for the fold change
353 calculated for the combined iDUX4 samples using a bootstrap approach. For a protein with n
354 peptides whose H/L ratios were used to calculate protein fold change, we calculated the
355 probability that n peptides chosen randomly out of all measured peptides would yield a similar
356 fold change by chance. To this end, we generated a null distribution for each protein as follows.
357 First, the same number of H/L ratios as the number of peptides quantified for that protein were
358 randomly chosen from the set of all calculated peptide-level H/L ratios (n=229,558). With these
359 randomly chosen H/L ratios, a fold change was calculated by computing the median. This process
360 was repeated 1,000 times to generate the null distribution of H/L ratios for that protein. Using this
361 null distribution, a *p*-value was calculated using a two-tailed Z test.

362

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

368

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

377 **Figure legends**

378

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

380 A) Schematic of experimental set up and the subsequent data analysis steps to measure protein
381 fold change in cells expressing vDUX4 or iDUX4.

382 B) Histogram of normalized, log2-transformed DUX4/Control ratios for all peptides mapping to
383 ZSCAN4, a DUX4 target gene (top panels), and RPL15, a housekeeping gene (bottom panels),
384 following 24 and 36 hours of vDUX4 expression.

385 C) Histogram of normalized, log2-transformed DUX4/Control ratios for all peptides mapping to
386 ZSCAN4, a DUX4 target gene (top panels), and RPL15, a housekeeping gene (bottom panels),
387 in the two label-swap replicates of iDUX4 expression.

388 D) Box plot of normalized, log2-transformed DUX4/Control ratio for peptides corresponding to the
389 FSHD biomarker transcripts identified by Yao et al (2015) in the vDUX4 dataset. Each dot refers
390 to an individual peptide that was quantified and the color represents the time point of vDUX4
391 expression (24 hours in blue and 36 hours in purple).

392 E) Box plot of normalized, log2-transformed DUX4/Control ratio for peptides corresponding to the
393 FSHD biomarker transcripts identified by Yao et al (2015) in vDUX4 dataset. Each dot refers to
394 an individual peptide that was quantified and the color denotes replicate 1 (red) or replicate 2
395 (purple) of iDUX4 expression.

396

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

399 A) Table showing number of genes for which we could quantify fold changes in RNA and protein
400 levels in the iDUX4 datasets.

401 B) Scatter plot of log_2 fold change in RNA levels (DUX4/control) versus log_2 fold change in protein
402 levels (DUX4/control) of all genes.
403 C) Gene Ontology analysis of genes that are up- or down-regulated > 4-fold at the RNA level or
404 protein level. Multiple testing correction done via Benjamini-Hochberg procedure.
405 D) Fluorescence micrographs of cells with and without DUX4 induction for 14h, stained with DAPI
406 (blue), anti-GM130 (Golgi apparatus; in green) and phalloidin-555 (actin; in red).
407 E) Quantification of percent of cells with DUX4-positive nuclei and fragmented Golgi apparatus.
408 Error bars represent standard deviation.
409
410 **Figure 3. Extensive post-transcriptional buffering of stress response genes in DUX4-**
411 **expressing cells.**
412 A) Scatter plot of log2 fold change of RNA levels (DUX4/control) versus log2 fold change of protein
413 levels (DUX4/control). Genes that show an upregulation at the RNA level, but no significant
414 change at the protein level fall in the blue rectangle.
415 B) RNA and protein fold change across the different datasets for representative genes that show
416 post-transcriptional buffering.
417 C) Changes in RNA levels of the chaperone HSPA5 over time of expression of vDUX4 expression
418 (or control) as measured by qRT-PCR. The region shaded in blue represents the region where
419 vDUX4 sample significantly deviates from Control due to induction of unfolded protein response.
420 D) Levels of phosphorylated eIF2α in control versus DUX4-expressing cells over an expression
421 time course as detected by immunoblots.
422 E) Percentage of [^{35}S]-Methionine incorporation (representing translation efficiency) in DUX4-
423 expressing cells normalized over control cells during a time course of DUX4 expression.
424
425 **Figure 4. DUX4 induces post-transcriptional gene regulation.**

426 A) Scatter plot of log2 fold change in RNA levels (DUX4/control) versus log2 fold change in
427 protein levels (DUX4/control). Genes regulated post-transcriptionally result in drastic changes
428 in protein level without major changes in RNA levels and are highlighted in gold.
429 B) Schematic representation of RNA- and protein-level changes for the genes involved in mRNA
430 surveillance. Colors represent a heat map of actual fold changes in RNA levels (top) and in
431 protein levels (bottom). Proteins with fewer than two quantified peptides are outlined by dotted
432 lines.
433 C) Scatter plot of log2 fold change in RNA levels (DUX4/control) versus log2 fold change in
434 protein levels (DUX4/control) for genes in the ubiquitin proteasome pathway.
435 D) Model for DUX4-induced post-transcriptional gene regulation.
436

437 **Supplementary Figures.**
438
439
440 **Figure S1. Demonstration of peptide to protein summarization for two candidate genes.**

441 A) The mass spectra for a representative peptide from a heavy-labeled sample.

442 B) Histogram of log2 transformed heavy to light ratio of peptides from cells labeled with heavy
443 amino acids for 1 week.

444 C) Histogram of log2 transformed heavy to light ratio of peptides from an equal mixture of
445 unlabeled cells with cells labeled with heavy amino acids for 1 week.

446 D) Histogram shows the DUX4/Control ratios (log2) for all peptides in the iDUX4 dataset (top
447 panel), peptides mapping to RPL15 (middle panel) or ZSCAN4 gene (bottom panel)

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

449

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

451 Screenshot of the Shiny web server showing a sample analysis of the RNA and protein-level data
452 for a DUX4 transcriptional target, ZSCAN4.

453

454

455 **Supplementary Data**

456 Supplementary data 1: PSM-level data (deposited to Dryad; Link:
457 https://datadryad.org/review?doi=doi:10.5061/dryad.ck06k75)

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

459 Supplementary file 3: R markdown document with the code needed to process mass spectrometry
460 data and generate figures.

461

462

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.

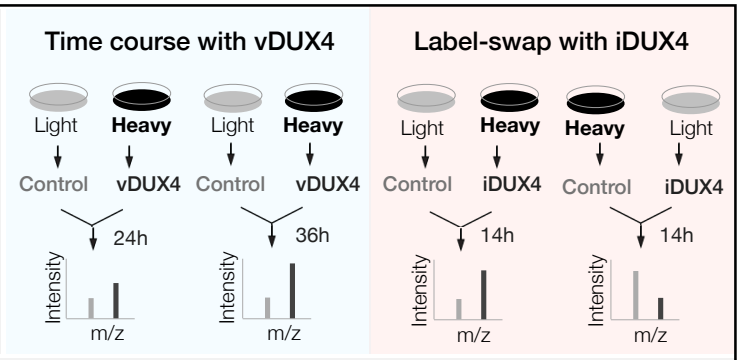
References

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. Jagannathan S*, et al.* (2016) Model systems of DUX4 expression recapitulate the transcriptional profile of FSHD cells. *Hum Mol Genet* 25(20):4419-4431.

530 21. Harsha HC, Molina H, & Pandey A (2008) Quantitative proteomics using stable isotope
531 labeling with amino acids in cell culture. *Nat Protoc* 3(3):505-516.
532 22. Yao Z*, et al.* (2014) DUX4-induced gene expression is the major molecular signature in
533 FSHD skeletal muscle. *Hum Mol Genet* 23(20):5342-5352.
534 23. Rodriguez-Boulan E & Musch A (2005) Protein sorting in the Golgi complex: shifting
535 paradigms. *Biochim Biophys Acta* 1744(3):455-464.
536 24. Bexiga MG & Simpson JC (2013) Human diseases associated with form and function of
537 the Golgi complex. *Int J Mol Sci* 14(9):18670-18681.
538 25. Claudio N, Dalet A, Gatti E, & Pierre P (2013) Mapping the crossroads of immune
539 activation and cellular stress response pathways. *EMBO J* 32(9):1214-1224.
540 26. Oslowski CM & Urano F (2011) Measuring ER stress and the unfolded protein response
541 using mammalian tissue culture system. *Methods Enzymol* 490:71-92.
542 27. Palacios IM (2013) Nonsense-mediated mRNA decay: from mechanistic insights to
543 impacts on human health. *Brief Funct Genomics* 12(1):25-36.
544 28. Licklider LJ, Thoreen CC, Peng J, & Gygi SP (2002) Automation of nanoscale
545 microcapillary liquid chromatography-tandem mass spectrometry with a vented column.
546 *Anal Chem* 74(13):3076-3083.
547 29. Kall L, Canterbury JD, Weston J, Noble WS, & MacCoss MJ (2007) Semi-supervised
548 learning for peptide identification from shotgun proteomics datasets. *Nat Methods*
549 4(11):923-925.
550 30. Park SS*, et al.* (2012) Effective correction of experimental errors in quantitative proteomics
551 using stable isotope labeling by amino acids in cell culture (SILAC). *J Proteomics*
552 75(12):3720-3732.
553 31. Ong SE & Mann M (2006) A practical recipe for stable isotope labeling by amino acids in
554 cell culture (SILAC). *Nat Protoc* 1(6):2650-2660.
555

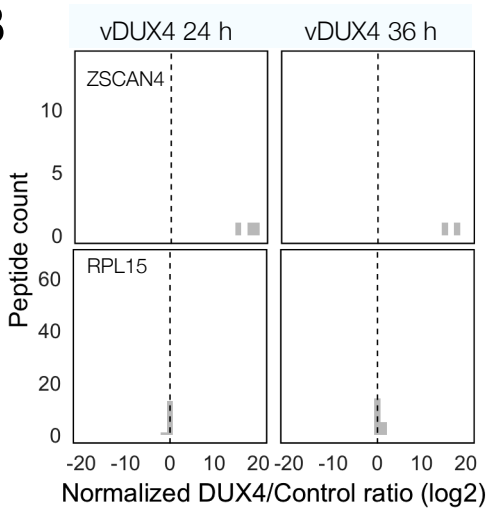
Figure 1

A

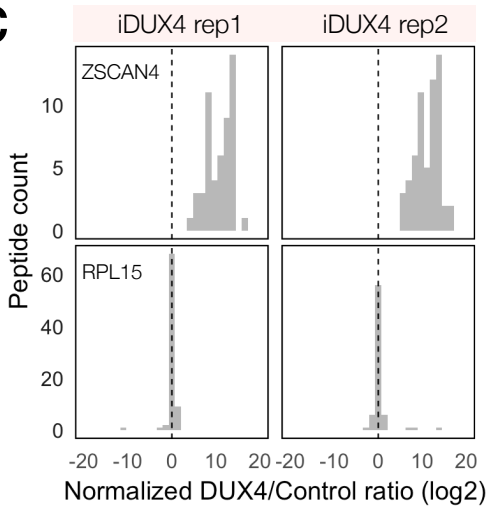


Experimental design	Time course with vDUX4		Label-swap with iDUX4	
Number of peptide-spectrum matches (PSMs)	114,143	112,814	249,986	253,592
Number of PSMs after filtering for quality	43,222	43,683	116,999	120,651
PSMs after filtering peptides with poor gene mapping	39,984	40,320	109,295	112,465
Summarize peptide to protein ratios via median; P values calculated via bootstrapping	2902 Proteins	2892 Proteins	4877 Proteins	4441 Proteins

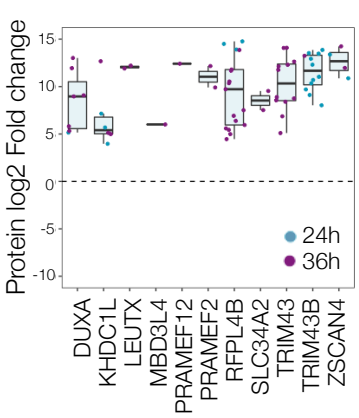
B



C



D



E

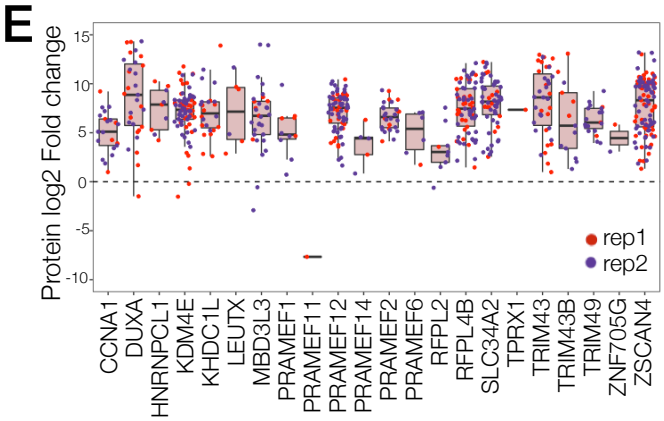
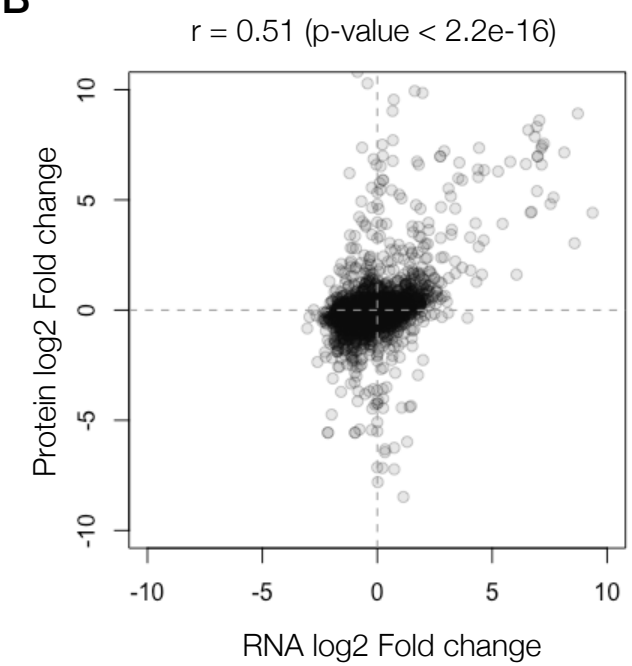


Figure 2

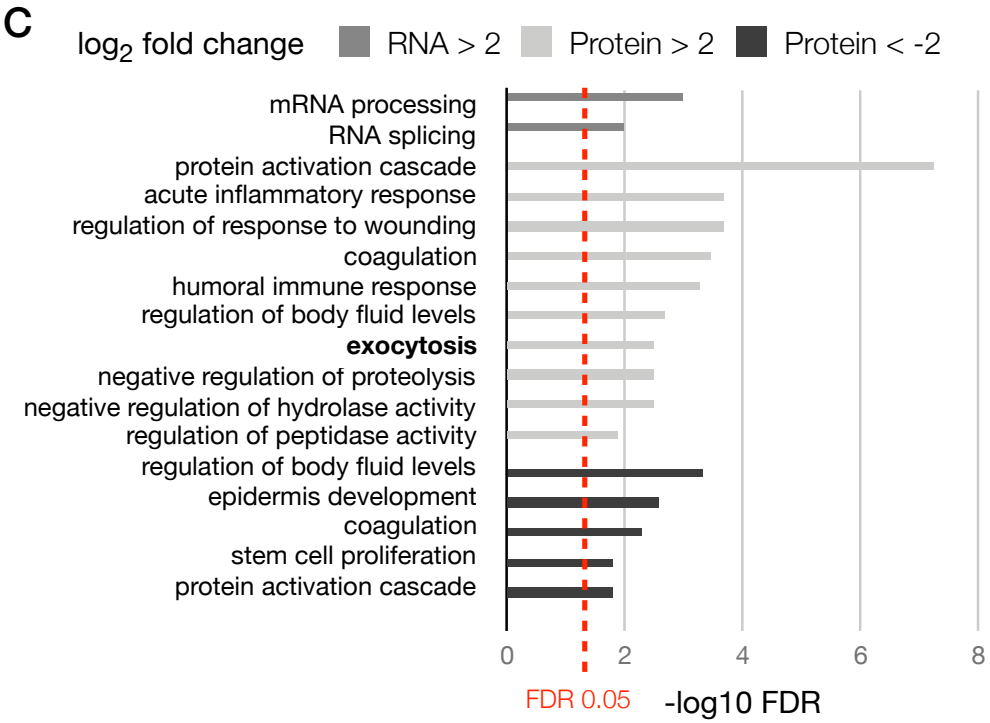
A

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

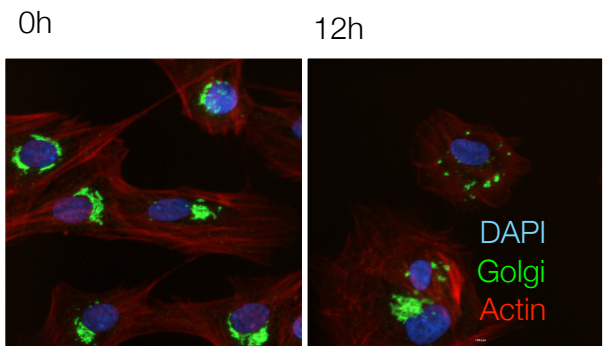
B



C



D



E

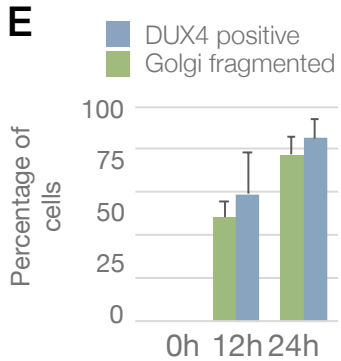


Figure 3

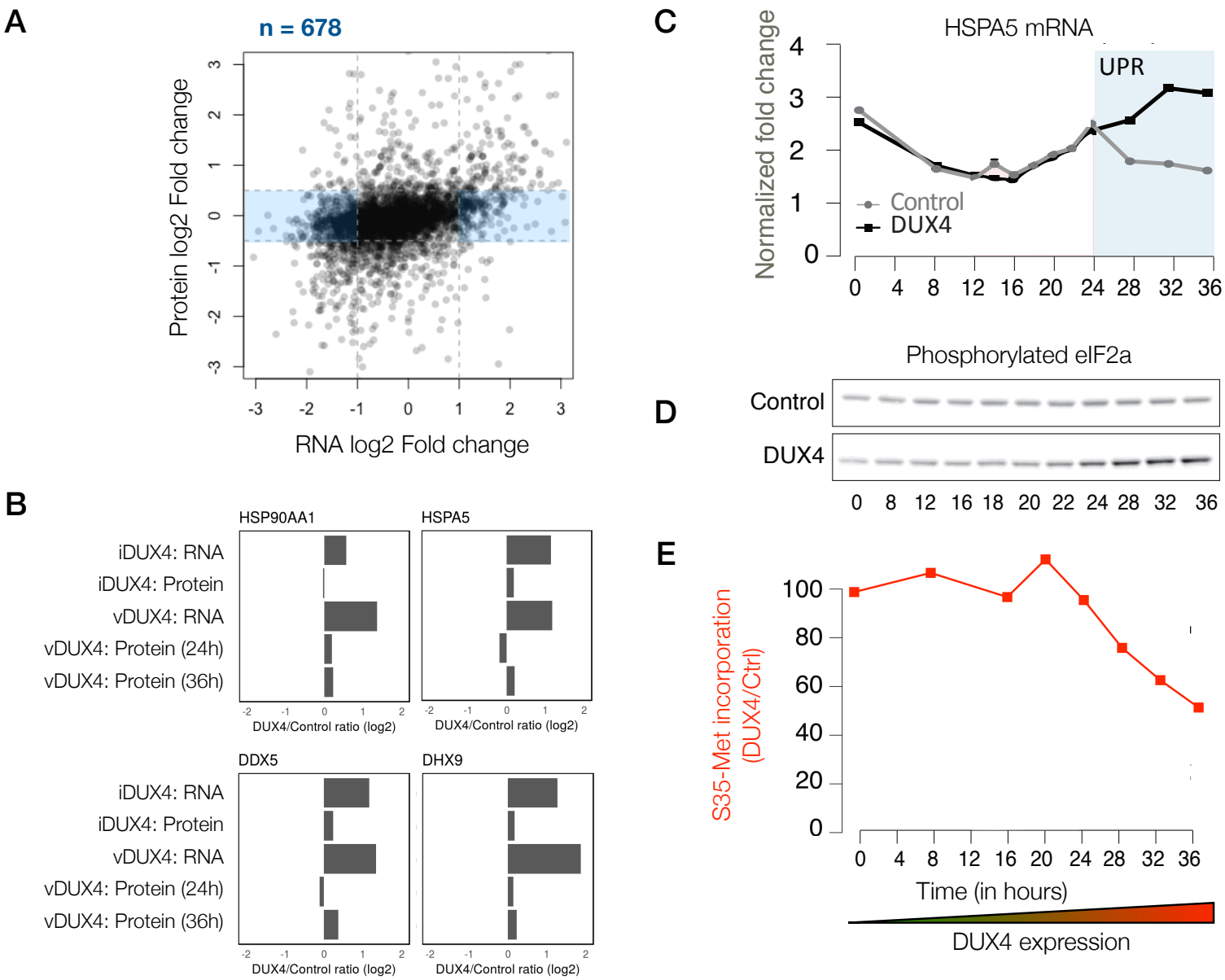

A
n = 678
Protein log2 Fold change
RNA log2 Fold change
B
HSP90AA1
HSPA5
DDX5
DHX9
iDUX4: RNA
iDUX4: Protein
vDUX4: RNA
vDUX4: Protein (24h)
vDUX4: Protein (36h)
DUX4/Control ratio (log2)
C
HSPA5 mRNA
UPR
Normalized fold change
Control
DUX4
D
Phosphorylated eIF2a
Control
DUX4
E
S35-Met incorporation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

Figure 4

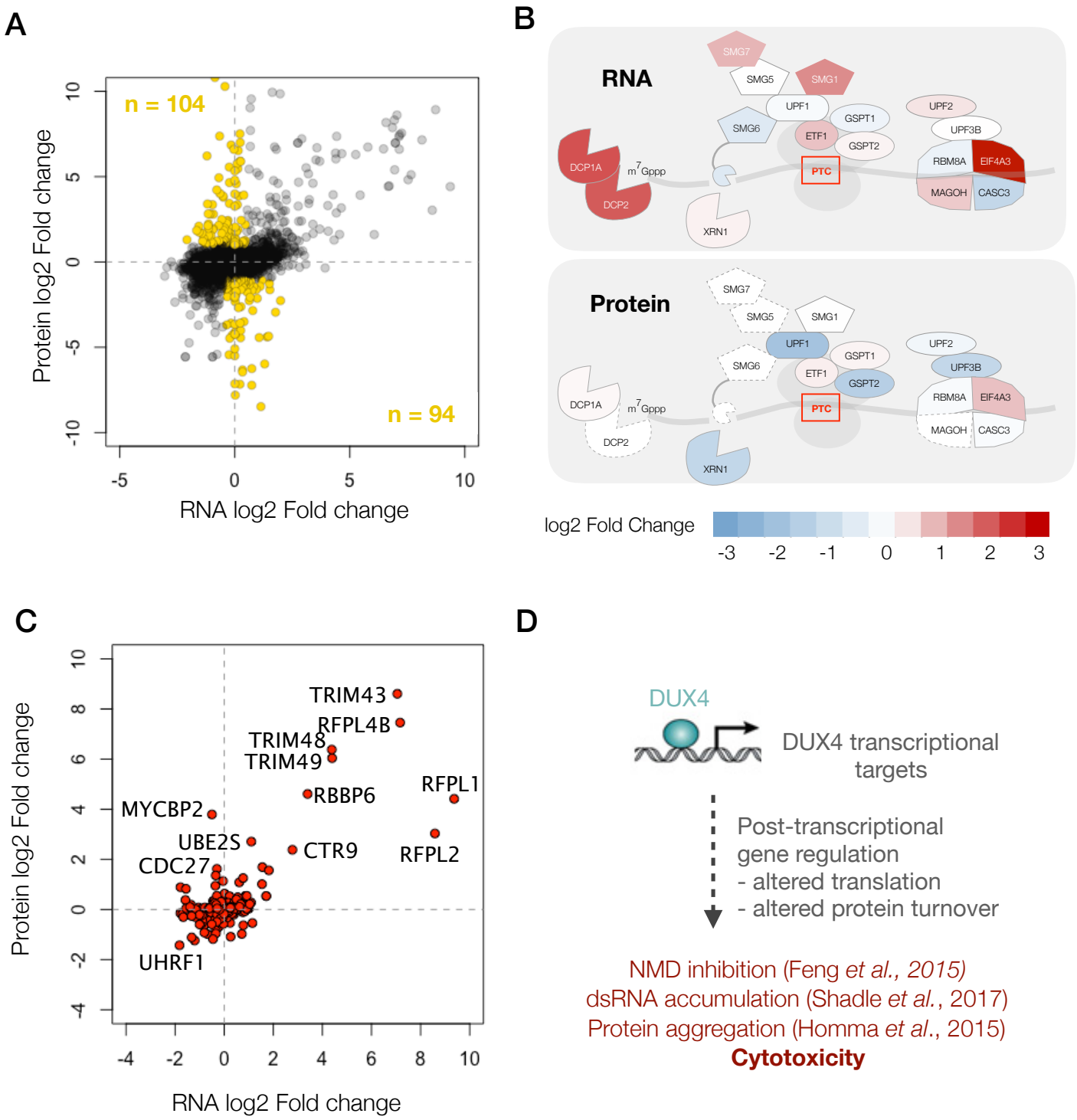

A
n = 104
n = 94
Protein log2 Fold change
RNA log2 Fold change
B
RNA
Protein
SMG7
SMG5
SMG1
UPF1
SMG6
GSPT1
ETF1
GSPT2
UPF2
UPF3B
RBM8A
EIF4A3
MAGOH
CASC3
DCP1A
DCP2
m7Gppp
PTC
XRN1
log2 Fold Change
C
TRIM43
RFPL4B
TRIM48
TRIM49
RBBP6
RFPL1
MYCBP2
UBE2S
CTR9
RFPL2
CDC27
UHRF1
Protein log2 Fold change
RNA log2 Fold change
D
DUX4
DUX4 transcriptional targets
Post-transcriptional gene regulation
- altered translation
- altered protein turnover
NMD inhibition (Feng et al., 2015)
dsRNA accumulation (Shadle et al., 2017)
Protein aggregation (Homma et al., 2015)
Cytotoxicity

Figure S1

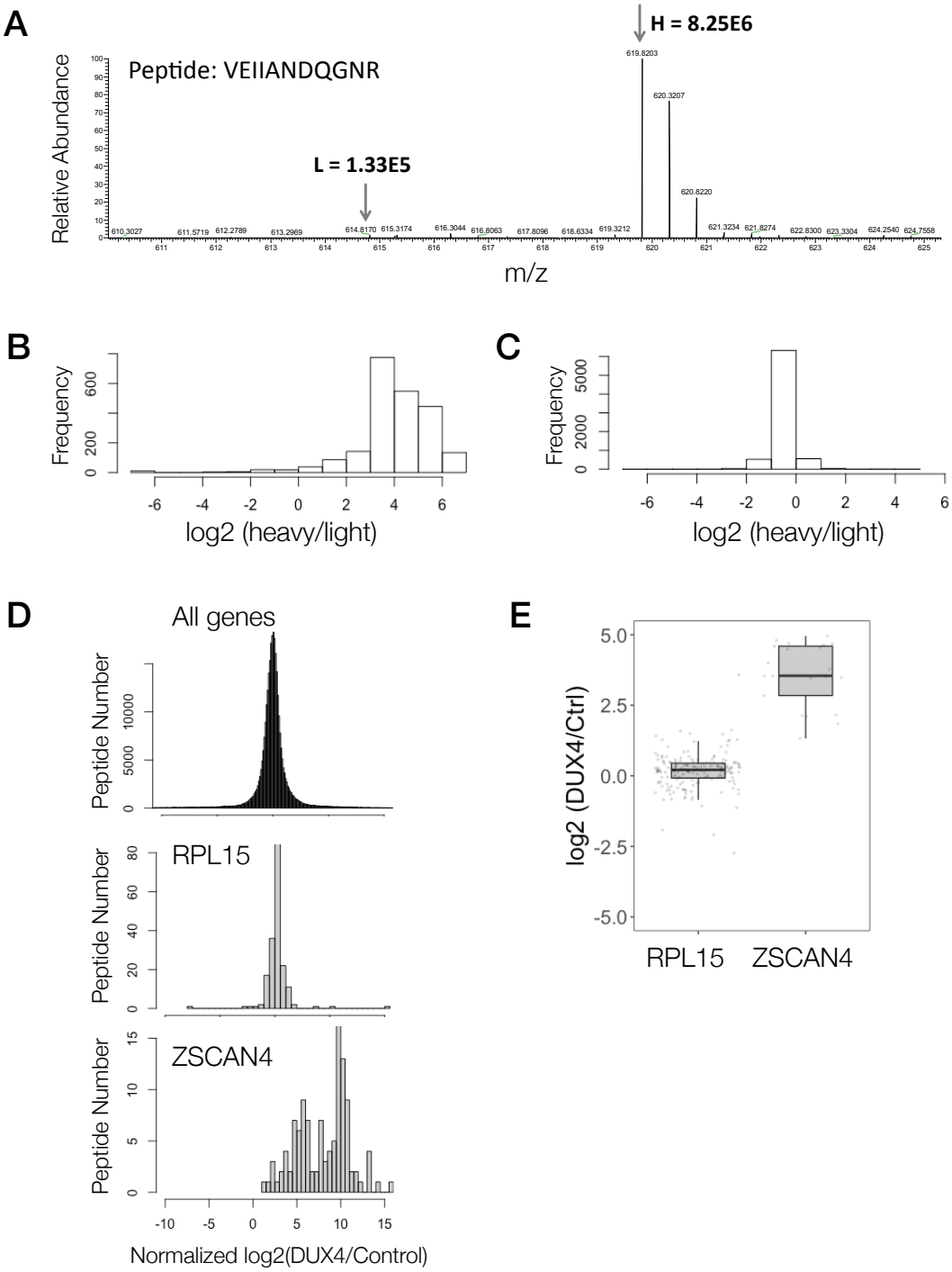
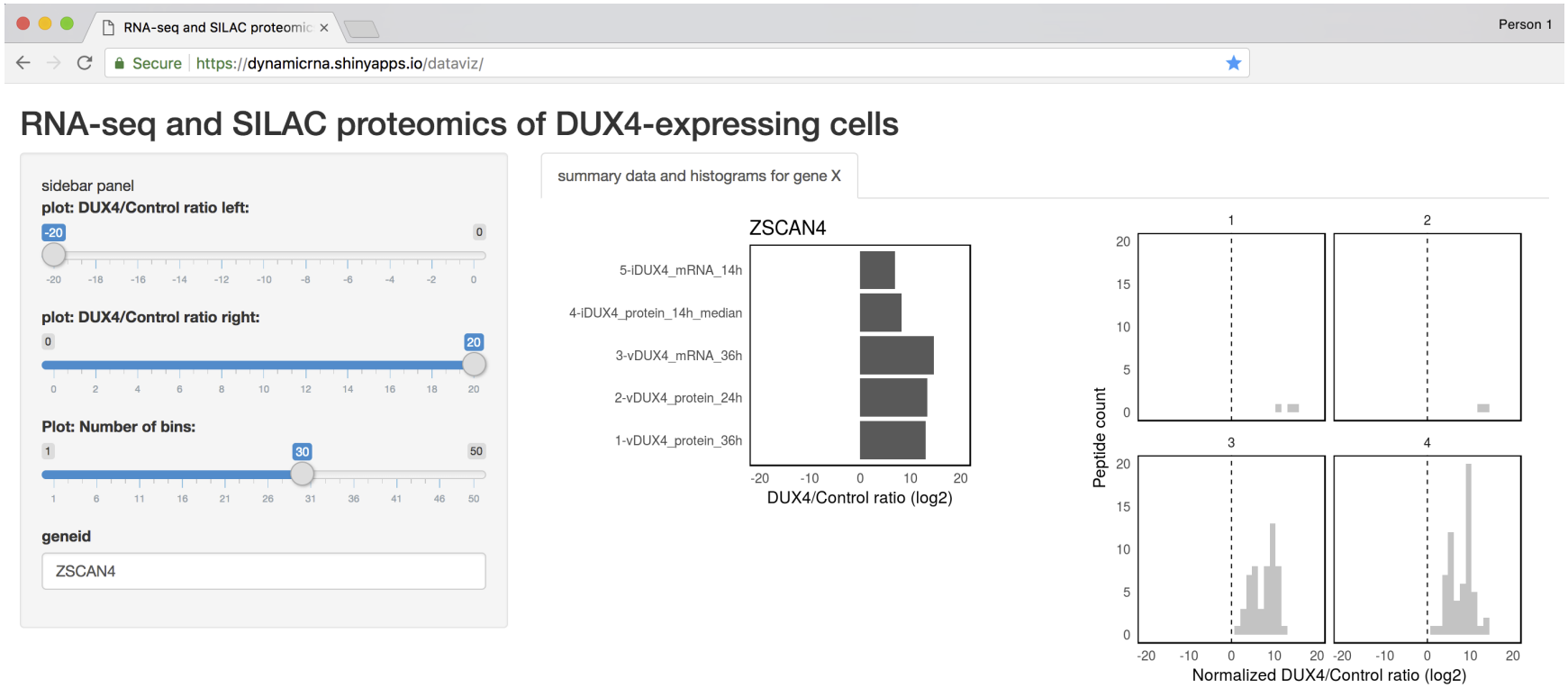

A
Peptide: VEIIANDQGNR
H = 8.25E6
L = 1.33E5
Relative Abundance
m/z
B
Frequency
log2 (heavy/light)
C
Frequency
log2 (heavy/light)
D
All genes
RPL15
ZSCAN4
Peptide Number
Normalized log2(DUX4/Control)
E
log2 (DUX4/Ctrl)
RPL15
ZSCAN4

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)