The unfolded protein response and its potential role in Huntington’s disease elucidated by a systems biology approach [version 1; peer review: 2 approved]

Ravi Kiran Reddy Kalathur¹, Joaquin Giner-Lamia¹, Susana Machado¹, Kameshwar R S Ayasolla¹, Matthias E. Futschik¹,²

¹Centre for Biomedical Research, University of Algarve, Faro, 8005-139, Portugal
²Centre of Marine Sciences, University of Algarve, Faro, 8005-139, Portugal

Abstract
Huntington’s disease (HD) is a progressive, neurodegenerative disease with a fatal outcome. Although the disease-causing gene (huntingtin) has been known for over 20 years, the exact mechanisms leading to neuronal cell death are still controversial. One potential mechanism contributing to the massive loss of neurons observed in the brain of HD patients could be the unfolded protein response (UPR) activated by accumulation of misfolded proteins in the endoplasmic reticulum (ER). As an adaptive response to counter-balance accumulation of un- or misfolded proteins, the UPR upregulates transcription of chaperones, temporarily attenuates new translation, and activates protein degradation via the proteasome. However, persistent ER stress and an activated UPR can also cause apoptotic cell death. Although different studies have indicated a role for the UPR in HD, the evidence remains inconclusive. Here, we present extensive bioinformatic analyses that revealed UPR activation in different experimental HD models based on transcriptomic data. Accordingly, we have identified 58 genes, including RAB5A, HMGB1, CTNNB1, DNM1, TUBB, TSG101, EEF2, DYNC1H1 and SLC12A5 that provide a potential link between UPR and HD. To further elucidate the potential role of UPR as a disease-relevant process, we examined its connection to apoptosis based on molecular interaction data, and identified a set of 40 genes including ADD1, HSP90B1, IKBKB, IKBKG, RPS3A and LMNB1, which seem to be at the crossroads between these two important cellular processes.

Keywords
Unfolded protein response (UPR), Huntington’s disease, Apoptosis, UPR interactome, HTT interactome
This article is included in the Rare diseases collection.

Corresponding author: Matthias E. Futschik (mfutschik@ualg.pt)

Competing interests: No competing interests were disclosed.

Grant information: The work presented was supported by CHDI Foundation (A-2666) and by the Portuguese Fundação para a Ciência e a Tecnologia (SFRH/BPD/70718/2010 to RKK and IF/00881/2013 to MEF).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2015 Kalathur RKR et al. This is an open access article distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the Creative Commons Zero "No rights reserved" data waiver (CC0 1.0 Public domain dedication).

How to cite this article: Kalathur RKR, Giner-Lamia J, Machado S et al. The unfolded protein response and its potential role in Huntington’s disease elucidated by a systems biology approach [version 1; peer review: 2 approved] F1000Research 2015, 4:103 (https://doi.org/10.12688/f1000research.6358.1)

First published: 01 May 2015, 4:103 (https://doi.org/10.12688/f1000research.6358.1)
Introduction

Huntington’s disease (HD) is an autosomal-dominant neurodegenerative disorder. Its symptoms include loss of motor control, cognitive decline, and behavioural abnormalities. In most cases, the onset of the disease occurs between the age of 35 and 50. The outcome is always fatal with a life expectancy following the disease onset of around 20 years.

The treatment of HD has remained symptomatic, as currently there is still no cure. The cause of HD is a mutation in a single gene called huntingtin (HTT). In HD patients, an expansion of the CAG repeat in exon 1 of huntingtin has been identified. This mutation results in an extended stretch of polyglutamine close to the N-terminus of the Huntingtin protein (HTT), which is involved in multiple molecular functions. Although the molecular cause has been known now for almost 20 years, the exact mechanisms leading to the observed massive cell death of neurons in the caudate nucleus of HD patients still await full clarification. A variety of processes such as excitotoxicity, protein aggregation and transcriptional dysregulation have been suggested to contribute to neurodegeneration in HD. More recently, several studies have indicated that the unfolded protein response (UPR) might be implicated in neurodegenerative diseases including HD.

ER stress and UPR

The endoplasmic reticulum (ER) is a crucial organelle for the correct folding and modification of proteins. Upon the accumulation of unfolded or misfolded protein in the ER, several transcriptional and translational mechanisms are triggered to ensure fidelity of protein folding. This stress response is better known as the UPR. In particular, the UPR stress sensors Inositol-requiring protein-1 (IRE1), activating transcription factor 6 (ATF-6) and PKR-like ER kinase (PERK) are activated in mammalian cells when the ER exceeds its capacity for correct folding. As an adaptive response to counter-balance accumulation of un- or misfolded proteins, the UPR upregulates transcription of chaperones, temporarily attenuates new translation, and activates protein degradation via the proteasome. The main function of the UPR is to re-establish homeostasis by increasing the overall folding capacity. Although the primary role of UPR is an adaptive one, persistent ER stress can mediate toxicity and eventually lead to apoptosis through activation of JNK, ASK1 and caspase-12. Figure 1 depicts the different mechanisms and outcomes of UPR activation.

UPR and HD

ER stress and UPR have been indicated for a variety of neurodegenerative disorders, where protein misfolding plays a significant role. For HD, finding a direct connection appears to be an enigma at first glance, since HTT is not located in the ER and thus its misfolding should not trigger UPR. However, several proposals have been put forward to describe how mutant HTT (mHTT) might provoke ER stress through interference with different processes such as vesicular transport or ER-associated degradation (ERAD) resulting in accumulation of (misfolded) protein in the ER.

Although various lines of investigations have shown a potential role for UPR in the pathogenesis of HD, it remains difficult to assess its overall influence, given that the animal models and cell lines used in each individual study display great variability and distinct characteristics. Furthermore, most studies addressing the connection between UPR and HD focus on a small set of genes and proteins for UPR in the pathogenesis of HD, it remains difficult to assess its overall influence, given that the animal models and cell lines used in each individual study display great variability and distinct characteristics. Furthermore, most studies addressing the connection between UPR and HD focus on a small set of genes and proteins. As the UPR presents a potentially important process in HD progression and a novel therapeutic target, we aimed to complement these previous studies with systematic and comprehensive bioinformatic analyses. Accordingly, using a systems biology approach, we gathered all available data and focused on detecting the activation of UPR during HD, and also elucidating the potential connection between UPR and apoptosis in HD.
First, we assembled different sets of genes associated with the UPR and examined whether the included genes show differential expression in HD models or patients, when compared to controls. Next, we examined the promoter regions of upregulated UPR genes and detected significant enrichment of characteristic stress response elements. Additionally, we performed functional enrichment analysis on differentially expressed genes and found major biological processes implicated in UPR to be significantly over-represented. Furthermore, we assembled the UPR interactome and identified common proteins involved with apoptotic processes as well as with known HTT interactors, since those could provide crucial links between apoptosis and HD.

Materials and methods
Derivation of gene sets for UPR, apoptosis and HD
Since the UPR is a complex process, it is challenging to define a unique set of associated genes. Accordingly, we compiled three alternative gene sets that are either directly or indirectly involved in UPR, gathered from three distinct sources. The first termed, UPR-KEGG-GO ($n=265$), was derived from Gene Ontology\textsuperscript{23} (http://geneontology.org/page/go-database) (RRID:nif-0000-20935) and Pathway\textsuperscript{24,25} (http://www.genome.jp/kegg/pathway.html) (RRID:nlx_31015) databases indicated in Table 1. The second, referred to as UPR-interaction network ($n=281$), was generated by assembling molecular interactions of UPR components ATF6, ATF4, DDIT3, EIFAK3, ERNI and XBP1 using UniHI\textsuperscript{26} (http://www.unihi.org/) (RRID:nif-0000-03609) and HDNetDB databases (http://hdnetdb.sysbiolab.eu). The third gene set, labelled UPR-literature ($n=2048$) was compiled from published experimental results\textsuperscript{27–30} performed in yeast and human cells using high-throughput techniques such as yeast two-hybrid, microarrays and ribosome profiling coupled with next generation sequencing, as well as from text-mining of the GeneCards database\textsuperscript{31} (http://www.genecards.org) (RRID:nif-0000-02879) (Table 1).

In order to examine the genes involved in the cross-talk between UPR and apoptosis, we derived a list of genes that are either directly or indirectly involved in apoptosis from several different sources.

### Table 1. Data sources used to compile UPR gene sets.

<table>
<thead>
<tr>
<th>Gene set</th>
<th>Sub-classification</th>
<th>Number of genes</th>
<th>Total number of unique genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>UPR-KEGG-GO</td>
<td>Pathway</td>
<td></td>
<td>265</td>
</tr>
<tr>
<td></td>
<td>KEGG:04141: Protein processing in ER</td>
<td>165</td>
<td></td>
</tr>
<tr>
<td></td>
<td>REACTOME: Unfolded protein response</td>
<td>79</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gene ontology</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GO:003043: ER-associated ubiquitin-dependent protein catabolic process</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GO:0030968: endoplasmic reticulum unfolded protein response</td>
<td>89</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GO:0034976: response to endoplasmic reticulum stress</td>
<td>142</td>
<td></td>
</tr>
<tr>
<td>UPR-interaction Network</td>
<td>Interactions</td>
<td></td>
<td>281</td>
</tr>
<tr>
<td></td>
<td>ATF4</td>
<td>92</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ATF6</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DDIT3</td>
<td>73</td>
<td></td>
</tr>
<tr>
<td></td>
<td>EIF2AK3</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ERN1</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td></td>
<td>XBP1</td>
<td>111</td>
<td></td>
</tr>
<tr>
<td>UPR Literature</td>
<td>Publications (High through-put experiments)</td>
<td></td>
<td>2048</td>
</tr>
<tr>
<td></td>
<td>Study</td>
<td>Species</td>
<td>Method</td>
</tr>
<tr>
<td></td>
<td>Labunsky VM \textit{et al.}\textsuperscript{27}</td>
<td>Yeast</td>
<td>Ribosome profiling coupled with NGS</td>
</tr>
<tr>
<td></td>
<td>Travers KJ \textit{et al.}\textsuperscript{28}</td>
<td>Yeast</td>
<td>DNA microarray</td>
</tr>
<tr>
<td></td>
<td>Jonikas MC \textit{et al.}\textsuperscript{29}</td>
<td>Yeast</td>
<td>Synthetic genetic array methodology &amp; High-throughput flow cytometry</td>
</tr>
<tr>
<td></td>
<td>Christianson JC \textit{et al.}\textsuperscript{30}</td>
<td>Human</td>
<td>Affinity purification, LC-MS/MS, High-throughput Y2H</td>
</tr>
<tr>
<td>GeneCards</td>
<td>GeneCards: ER-Stress</td>
<td>275</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GeneCards: Unfolded Protein Response</td>
<td>325</td>
<td></td>
</tr>
</tbody>
</table>
sources, namely the Gene Ontology database (GO:0006915; Apoptotic process; n=431), KEGG pathway (hsa04210; Apoptosis; n=88), Reactome pathway database (REACT_578.8; Apoptosis; n=148; http://www.reactome.org/) (RRID:nif-0000-03390) and literature reviews (n=85) (ii)-c. All genes included were annotated to be involved in the induction of apoptosis, anti-apoptosis, regulation of apoptosis or were caspases (including both activators and inhibitors).

For establishing putative links to HD, we additionally put together two other gene sets:

(i) HD therapeutic targets (HDTT) comprising 1033 genes. This set includes genes which were annotated by the curators of the HD Research CrossRoads database as being associated with HD based on experimental evidence, making them potential therapeutic targets. A detailed description of this gene set is provided elsewhere by Kalathur et al. (iii). The list of HDTT can be accessed at http://hdtt.sysbiolab.eu/.

(ii) HTT interactors (HTT-int) including 1015 genes whose proteins have been shown to interact, or to be physically associated with HTT based on a diverse range of experiments. The large number of interactors can be explained through the inclusion of high-throughput affinity purification experiments, which frequently results in the addition of indirect interactions (e.g. within complexes). This set of interactors was obtained from the HDNetDB database (http://hdnetdb.sysbiolab.eu/).

Collection and processing of HD gene expression data
All HD gene expression data used for this study were downloaded from the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/) (RRID:nif-0000-00142). These data include gene expression from human brain and blood samples, human iPSCs, mouse, rat and yeast HD models, as well as murine cell lines (Table 2). All expression data sets were pre-processed using RMA (Robust Multi-array Average) implemented in R (available at http://www.r-project.org/) (RRID:nif-0000-10474) and analysed using several Bioconductor packages (iii). To enable the comparison across organisms, we mapped genes from mouse, rat, worm and yeast to orthologous human genes using HUGO Comparison of Orthology Predictions (HCOP) search tool (available online at http://www.genenames.org/cgi-bin/hcops), which is based on integrated data from HUGO Gene Nomenclature Committee (HGNC) (iv) (RRID:nif-0000-02955).

Identification of differential expression using Gene Set Enrichment Analysis (GSEA)
We performed gene set enrichment analysis (GSEA) (v) (RRID:nif-0000-30629) comparing HD-associated expression to wild type or control data to identify differentially expressed genes. As input, we used the above-mentioned UPR gene sets and HD gene expression data. UPR genes were identified as significant when the enrichment score (ES) corresponded to a false discovery rate (fdr) ≤ 0.05 in HD gene expression data sets. For further analysis, we used only the genes present in the ‘UPR core enrichment’ gene sets. Those genes belonged to the leading-edge subsets and contributed the most to the enrichment scores, and are the most differentially expressed among the UPR genes. To visualize these results, we generated Venn diagrams using jvenn (vi), to display the common genes across alternative comparisons.

Identification of stress response elements in the promoter regions
In order to verify the presence of unfolded protein response element (UPRE) and ER stress response elements (ERSE I and II) (vii) in the upstream regions (-1000bp to +500bp) of UPR genes upregulated in HD, we downloaded all the human promoter regions, (n=23322) available in the eukaryotic promoter database (EDP; http://epd.vital-it.ch) (viii) (RRID:nif-0000-02806). Next, we used Regulatory Sequence Analysis Tools (RSAT) (ix) to map these elements in the promoters and computed the enrichment of these stress elements in promoters of upregulated UPR genes compared to all the human promoters using hypergeometric test (equivalent to Fisher’s exact test).

Table 2. List of HD gene expression datasets used for the gene set enrichment analysis.

<table>
<thead>
<tr>
<th>GEO ID</th>
<th>Sample</th>
<th>Organism</th>
<th>Pubmed id</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSE3790</td>
<td>HD (CN) vs Control</td>
<td>Human</td>
<td>16467349</td>
</tr>
<tr>
<td>GSE24250</td>
<td>HD vs Control (blood)</td>
<td>Human</td>
<td>21969577</td>
</tr>
<tr>
<td>GSE37547</td>
<td>HD-iPSc vs corrected HD iPSc</td>
<td>Human</td>
<td>22748967</td>
</tr>
<tr>
<td>GSE3821</td>
<td>R6/1-18w, 22w, 27w vs WT</td>
<td>Mouse</td>
<td>17696994</td>
</tr>
<tr>
<td>GSE9803</td>
<td>R6/2-12w vs WT</td>
<td>Mouse</td>
<td>17519223</td>
</tr>
<tr>
<td>GSE10202</td>
<td>CHL2-22m vs WT</td>
<td>Mouse</td>
<td>17519223</td>
</tr>
<tr>
<td>GSE9330</td>
<td>Ctip2-KO vs WT</td>
<td>Mouse</td>
<td>18199763</td>
</tr>
<tr>
<td>GSE18551</td>
<td>YAC128-12, 24m vs WT</td>
<td>Mouse</td>
<td>20089533</td>
</tr>
<tr>
<td>GSE3583</td>
<td>HdhQ111 vs WT</td>
<td>Mouse</td>
<td>17708681</td>
</tr>
<tr>
<td>GSE9760</td>
<td>mESC (CAG150)-d4, d6 vs WT</td>
<td>Mouse</td>
<td>NA</td>
</tr>
<tr>
<td>GSE12481</td>
<td>Neuronal-culture 82Q vs CT</td>
<td>Rat</td>
<td>18815258</td>
</tr>
<tr>
<td>GSE18644</td>
<td>Htt103Q vs Htt25Q</td>
<td>Yeast</td>
<td>21044956</td>
</tr>
</tbody>
</table>
Functional enrichment analysis
To identify enriched biological processes in our gene sets we used BiNGo\textsuperscript{16} (RRID:nlx_149196) for Cytoscape\textsuperscript{20} (http://apps.cytoscape.org/apps/bingo) (RRID:nif-0000-30404); and GSEA\textsuperscript{22} (RRID:nif-0000-30629) to evaluate if genes from curated Reactome pathways (obtained from the Molecular Signature Database (MsigDB)\textsuperscript{23}) were statistically over-represented. The significance of each identified biological process or pathway was calculated using the hypergeometric test, adjusted for multiple testing and converted to \textit{fdr} using the Benjamini and Hochberg method\textsuperscript{24} implemented in BiNGo (RRID:nlx_149196) or in GSEA (RRID:nif-0000-30629), respectively. We considered only those processes and pathways with an \textit{fdr} of \leq 0.05 to be significantly enriched.

Results and discussion
Identification of common UPR genes across HD gene expression studies
To determine possible implications of the UPR in HD, we sought to assess its activation using a computational approach and the evaluation of existing data. First, we catalogued genes that are involved in UPR from several different sources and divided them into three different categories: UPR-KEGG-GO, UPR-interactions and UPR-literature as described in Material and Methods section and detailed in Table 1 and Supplementary Figure 1.

Since changes in gene transcription are main effects of UPR activation and published microarray data are available for HD in humans as well as for HD models, we collected 12 different gene expression datasets generated for the study of HD: three datasets included expression from human blood and brain samples as well as human induced pluripotent stem cells (iPSCs); seven datasets were derived from HD mouse models and cell cultures; one from rat cells and one from yeast cells. If the expression dataset constituted time-series, we split the dataset according to the time points to maintain the temporal aspect of the expression changes.

We reasoned that UPR activity should be reflected in the regulation of UPR genes. By applying GSEA we tested whether UPR genes tend to be differentially expressed in HD samples or models compared to the corresponding controls. GSEA was employed, since it is able to detect modest but consistent tendencies in expression datasets generated for the study of HD: three datasets included expression from human blood and brain samples as well as human induced pluripotent stem cells (iPSCs); seven datasets were derived from HD mouse models and cell cultures; one from rat cells and one from yeast cells. If the expression dataset constituted time-series, we split the dataset according to the time points to maintain the temporal aspect of the expression changes.

For each of the compiled UPR genes sets, the normalized enrichment scores (NES) produced by GSEA are shown for different comparisons of HD-associated expression with controls. Positive scores indicate a tendency towards upregulation; negatives scores indicate a tendency towards downregulation of genes in the UPR sets. Comparisons that showed significant upregulation of UPR gene sets (\textit{fdr} \leq 0.05 and NES \geq +1.4) are highlighted by red background, while significant downregulation (\textit{fdr} \leq 0.05 and NES \leq +1.4) by green background.
that were commonly upregulated, when UPR activation was indicated (Figure 3, Supplementary data file 1). We refer to this set of genes as \textit{UPR-HD} sup. As UPR has been also associated with the suppression or degradation of a substantial number of genes\cite{1,2}, we carried out the equivalent procedure to identify UPR genes whose downregulation is conserved in different HD models. Here, we intersected the enrichments core from comparisons displaying suppression of UPR genes (Supplementary Figure 2, Supplementary data file 2). This resulted in 81 commonly downregulated genes. We refer to the combined set of UPR genes (consisting of 132 upregulated and 81 downregulated genes) that were commonly detected as differentially regulated in HD gene expression data as \textit{UPR-HD} diff.

Examining stress response elements in the promoter regions of upregulated UPR genes

To assess whether the observed upregulation indeed reflects the activation of UPR or if it might be a consequence of other unrelated processes, we carried out an analysis of the promoter regions of genes included in UPR-HD sup. We searched for the presence of sequence elements that indicate binding of transcription factors associated with the UPR. In particular, we searched for unfolded protein response elements (UPRE; TGACGTG (G/A)) and the alternative ER stress response elements I (ERSE-I; CCAAT(N9)CCACG) and II (ERSE-II; ATTGG-N-CCACG) in promoter regions (-1000 bp to +500 bp) regions of UPR-HD sup genes. These characteristic sequence elements are targeted by the bZIP transcription factors ATF6 and XBP1, which are main mediators of the transcriptional adaptation evoked by UPR\cite{3}.

Strikingly, we found that the vast majority of the UPR-HD sup has such a characteristic binding sequence in their promoter regions (Supplementary Figure 3). Compared to number of sequence elements that we would expect by chance, a highly significant overrepresentation was detected for the UPR-HD sup genes. More specifically, we found the occurrence of UPRE in 104 genes ($p=3.0\cdot10^{-9}$), ERSE-I in 93 genes ($p=0.0019$), and ERSE-II in 8 ($p=0.052$). Notably, a large number of UPR-HD sup had alternative binding motifs included in the promoter region: 70 genes had both ERSE-I and UPRE, 2 genes had both ERSE-I and ERSE-II and six genes had all three elements (for list of genes see Supplementary data file 3) which might suggest that these genes are under particularly tight control of UPR-associated transcription factors ATF6 and XBP1. Altogether, the results of the promoter analysis support the conclusion that the upregulation of UPR genes in HD models faithfully reflects an activated UPR.

**Figure 3.** Upregulated UPR genes. Genes included in the core enrichment set for comparisons that indicated UPR activation (highlighted in red in Figure 2) were compared. Common upregulated UPR genes (n=169) in five HD mouse models (left side) were intersected with upregulated UPR genes in human HD iPSCs (right side) resulting in a set of 132 UPR genes, whose activation was conserved across the different HD models. The bar plots (bottom) display the number of UPR genes that were assigned to the core enrichment sets for comparisons that indicated upregulation.
Biological processes that are enriched in differentially expressed UPR genes

Since the UPR comprises a complex series of diverse molecular mechanisms, we examined the functional composition of UPR-HD\(^\text{upreg}\) genes. For this purpose, we performed functional analysis using BiNGO to identify enriched biological processes (as defined in GO) that are overrepresented among UPR-HD\(^\text{upreg}\) genes. All the biological processes that are significantly enriched in our analysis are listed in Supplementary data file 4. Expectedly, we detected that stress-related functional categories such as ‘response to stress’ (GO ID:6950; \(n=44; fdr=2.08E-03\)) and ‘response to unfolded protein’ (GO ID:6986; \(n=13; fdr=4.55E-09\)) were enriched (Figure 4a). A second group of significantly overrepresented GO categories were related to ‘protein transport’ (GO ID:15031; \(n=34; fdr=1.76E-07\)) and ‘protein localization’ (GO ID:8104; \(n=36; fdr=1.19E-06\)) including ‘vesicle-mediated transport’ (GO ID:16192; \(n=24; fdr=9.03E-05\)) and ‘ER to Golgi vesicle-mediated transport’ (GO ID:6888; \(n=4; fdr=4.00E-02\)) (Figure 4b). Additionally, we found ‘ER-nucleus signalling pathway’ (GO ID:6984; \(n=9; \text{adjp-value}=1.81E-07\)) to be highly enriched. It has been previously reported that ER-nucleus signalling pathway functions via activation of NF-\(\kappa\)B due to ER-overload triggered by protein congestion\(^8\).

Furthermore, UPR-HD\(^\text{upreg}\) genes tended to be associated with protein catabolism and in particular protein degradation (Figure 4c). Significant processes here were e.g. ‘protein catabolic process’ (GO ID:30163; \(n=18; fdr=6.40E-05\)), ‘proteasomal ubiquitin-dependent protein catabolic process’ (GO ID:43161; \(n=13; fdr=6.69E-06\)) and ‘protein ubiquitination’ (GO ID:31396; \(n=7; fdr=3.12E-02\)). These results coincide well with previous studies establishing the connection of UPR and ERAD and showing, for instance, that the extent of activation of the UPR is concurrent with the severity of ERAD defect\(^6\).

Finally, genes linked to apoptosis could be found among the UPR-HD\(^\text{upreg}\) genes (Figure 4d). Of particular interest for potential intervention could be genes associated with ‘regulation of apoptosis’ (GO ID:42981; \(n=24; fdr=1.65E-02\)), as their manipulation may prevent the execution of the apoptotic pathway under persistent ER stress.

In summary, UPR genes detected as commonly differentially regulated in HD expression data were not restricted to a particular functional category, but can be associated with many processes linked to the UPR.

Pathways enriched in upregulated genes

Complementary to the functional composition, we evaluated whether specific pathways might be activated based on the observed commonly upregulated UPR genes (UPR-HD\(^\text{upreg}\)). Therefore, we carried out pathway enrichment analysis using a set of pathways curated in the Reactome database. As expected, ‘unfolded protein response’ (\(n=6; fdr=2.04E-05\)), ‘activation of genes by ATF4’ (\(n=3; fdr=0.00291\)) and ‘PERK regulated gene expression’ (\(n=3; fdr=0.00333\)) were detected as significantly enriched among UPR-HD\(^\text{upreg}\) genes (Figure 5). More interestingly, we also found an overrepresentation of components of the ‘immune system’ (\(n=14; fdr=6.74E-05\), ‘adaptive immune system’ (\(n=6; fdr=1.96E-03\)), ‘NGF signalling’ (\(n=9; fdr=2.42E-03\)), and ‘Diabetes pathways’ (\(n=7; fdr=2.04E-05\)). Complete results from the analysis are included in Supplementary data file 5.

Remarkably, recent studies have also suggested that ER stress and activated UPR are interconnected with inflammatory processes\(^6\). Inflammation is an immunological process usually carried out by the vascular system to counteract disease, and to fight foreign antigens against invasion. Within the brain, microglia and astrocytes play important immunological functions. Until very recently, little was known about inflammatory molecules in HD. Recent studies, however revealed a distinct profile of inflammatory mediators from post-mortem human HD tissue\(^6\). Inflammatory mediators such as IL-1\(\beta\) and TNF-\(\alpha\) were increased only in the striatum, whereas IL-6, IL-8 and MMP-9 were also upregulated in cortex and in the cerebellum\(^6\). This supports the conjecture that secreted inflammatory cytokines and activated microglia cells could lead to axonal damage and extensive neuronal cell death in HD pathology\(^6\).

In general, activated microglia exert their diverse effects on neurons and macroglia (astrocytes and oligodendrocytes). Inflammation occurs through the release of cytoprotective agents such as growth factors, plasminogen, and neuroprotective cytokine as well as cytokotoxic substances such as oxygen radicals, nitricoxide, glutamate, proteases, and neurotoxic cytokines. One of the earliest reports describing microglial abnormalities in HD was provided by Singhrao et al.\(^6\). Microglial cell counts were considerably increased in the caudate putamen of HD and these microglial cells expressed increased amounts of complement factors. A more detailed investigation of microglial morphological changes associated with HD was performed by Sapp et al.\(^6\). The authors localized morphologically activated microglial cells in the neostriatum, cortex and globus pallidus as well as in adjoining white matter of HD brains. Additionally, positron emission tomography (PET) studies using the ligand for benzodiazepine receptor (PK-11195), which labels activated microglia have been employed to study neuroinflammation. Using this technique, Tai et al.\(^6\) demonstrated that microglial activation in HD patients correlates with disease progression as assessed by loss of dopamine D2 receptor binding sites. Interestingly, Tai et al. could also demonstrate that microglial activation and release of cytokine IL-6 is observed in presymptomatic HD gene carriers and can be detected up to 15 years before predicted age of onset. These findings indicate the microglial inflammatory activation is an early event associated with subclinical progression of HD and may constitute a target for early therapeutic intervention.

Besides the indication of processes related to the immune response, results of the pathway enrichment analysis also pointed to diabetes. It has been shown that diabetes in Wolcott-Rallison syndrome (a rare autosomal recessive form of juvenile diabetes) is a result of high levels of ER stress caused by mutations in the PERK gene in pancreatic \(\beta\)-cells. In addition, studies have shown that HD patients show increased incidence of diabetes\(^6,65\) and HD transgenic mice develop hyperglycemia\(^6,67\). More recently it has been experimentally validated that HD transgenic mice develop intraneuronal inclusions in the pancreatic \(\beta\)-cells, causing an intrinsic defect in insulin production\(^6\).
Figure 4. Biological processes enriched among differentially regulated UPR genes. GO hierarchies for biological processes overrepresented in the set of differentially regulated UPR are shown. Nodes indicate specific GO terms and their size represents the number of included UPR genes. The significance of overrepresentation (enrichment) is visualized by colour-coding from yellow to orange with the latter representing higher significance. No colour indicates that the process is not significantly enriched ($\text{fdr} \geq 0.05$). The overrepresented biological processes can be split into four major themes: (a) response to stress, (b) protein transport, (c) protein catabolic process and (d) apoptosis.
Prioritization of UPR-HD connectors through integrative analysis

To narrow down the list of UPR-HD genes for further inspection, we utilized additional information, including a reference set of potential molecular targets for HD therapy that were made available through the HD Research Crossroads database initiated by the CHDI Foundation (see Kalathur et al. 2012). Genes were included by experts in the field based on the evaluation of published literature and in-house screens using a set of defined criteria (see Kalathur et al. 2012). For instance, a gene was considered as a potential HD TT if genetic or pharmacologic modification of its activity led to a change of a HD-related phenotype in a validated cell culture or organism model of HD. At present, this curated reference set constitutes the most comprehensive collection of HD TTs. In addition, we extracted genes, whose corresponding proteins were reported to be physically associated with HTT, from the HDNetDB database. We recently demonstrated that HTT interactors tend to be enriched in proteins that influence the toxicity of mHTT, and provide favourable candidates for the identification of molecular modifiers of HD.

Notably, the search for stress response elements in the upstream regions of these 13 genes revealed that eight genes (RAB5A, HMGB1, CTNNB1, DNM1, TCP1, TUBB, TSG101 and DNAJB1) possess either UPRE or ERSE or both elements in their promoters, suggesting that these genes are under direct control of UPR-associated transcription factors (Table 3).

Inspection of the genes possessing UPRE or ERSE elements in their promoter regions revealed that four of them (TCP1, CCT2, DNAJB1 and HMGB1) have been reported to act as chaperones. Besides being essential components of the UPR, molecular chaperones can modulate the aggregation and toxicity of proteins, including mHTT. TCP1 (CCT1) and CCT2 are components of the TCP1 ring complex (TRiC) that uses cycles of ATP-binding and hydrolysis to bind unfolded polypeptides and facilitate their folding. Notably, TRiC has been identified as a potent suppressor of mHTT mediated toxicity and inhibitor of the mHTT protein aggregation in vitro and in vivo. DNAJB1 belongs to the group of DnaJ/Hsp40 (Heat shock protein 40) proteins that are involved in protein translation, folding and translocation through regulating ATPase activity of the Hsp70s chaperones. In a PC12 cell model, experiments indicated that DNAJB1 attaches to soluble mHTT oligomers and recruits Hsp70 suppressing mHTT mediated toxicity. Finally, HMGB1 encodes for the High-mobility group box 1 protein (HMGB1), which has recently been demonstrated to have chaperone-like activity, inhibiting aggregation of various proteins. Overexpression of HMGB1 can also decrease the aggregation induced by extended polyQ stretches.
Table 3. Presence of ER stress-associated sequence motifs in the promoter regions of differentially regulated UPR genes that interact with HTT and were classified as potential HD therapeutic targets (HDTT). + indicates the presence of particular stress response element in the promoter regions (+1000 to -500 bp).

<table>
<thead>
<tr>
<th>Gene</th>
<th>UPRE</th>
<th>ERSE-I</th>
<th>ERSE-II</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAB5A</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>HMGBl</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>CTNNB1</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNBl</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>TCP1</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>TUBB</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>TSG101</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>DNAJB1</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>CCT2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EEF2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DYNC1H1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HSPA5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SLC12A5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Linking the UPR network to apoptosis and HTT

A crucial aspect of the UPR in the context of HD is the possibility that it can trigger apoptosis upon persistent ER stress. To obtain a comprehensive view of the connections between UPR and apoptosis, we applied a network approach. First, we generated the UPR interactome from known protein interactions of UPR core components, which we extracted from the UniHI and HDNetDB databases (Supplementary data file 6). Second, we compiled a list of genes (n=594) associated with apoptosis from several different sources (as described in the materials and methods). We then used this list to identify 40 proteins associated with apoptosis within the UPR interactome (Supplementary data file 7). These genes included, among others, Apoptosis Signal Regulating Kinase 1 (ASK1, also known as MAP3K5), whose knockout in primary neuron provided protection from ER stress-induced JNK activation and cell death triggered by polyQ fragments.

As the mutation in HTT can perturb the function of interacting proteins by aberrant binding, we checked for each of the 40 proteins whether they have been reported to physically associate with HTT. Using molecular interaction data collected in HDNetDB, we detected that six of the 40 proteins interact with HTT i.e. ADD1, HSP90B1, IKBKB, RPS3A, IKBKG and LMNB1 (Supplementary data file 7). A visualisation of the UPR interactome with apoptosis-related proteins and HTT interactors highlighted can be found in Figure 6.

Literature review showed that the two proteins kinases IKBKB and IKBKG, the laminin LMNB1 and the ribosomal protein RPS3A have been previously linked to neurodegenerative diseases. IKBKB and IKBKG are subunits of IkB kinase (IKK). They activate members of the NF-κB transcription factor family by phosphorylation of their inhibitor (IkB) leading to ubiquitination and destruction of IkB, thereby allowing activation of the NF-κB complex. NF-κB maintains the balance between cell survival and apoptosis. Although unrelated to ER stress, it has been shown that inhibition of IKBKB decreases HTT proteolysis in a cell model, and thus might lower the load of toxic HTT fragments in HD. Recently, it has been reported that ubiquitination of IKBKG by Parkin, an ubiquitin ligase associated with Parkinson’s disease regulates the anti-apoptotic pathway that is key to maintaining mitochondrial integrity.

Lamin B1 protein, LMNB1 is thought to be involved in nuclear stability and chromatin structure. Experiments in Caenorhabditis elegans overexpressing aggregation-prone peptides identified laminins as modulators of protein toxicity at neuromuscular junctions. Further, in leukodystrophy mouse models, lamin B1 acts as an important regulator of myelin formation and maintenance, in humans lamin B1 gene duplications and large deletions upstream of promoter regions can cause autosomal-dominant leukodystrophies. More importantly, a recent study reports increased levels of lamin B1 in both human HD patients and the R6/1 mouse model of HD. Due to the involvement of lamin B1 in several cellular alterations such as chromatin organisation, gene transcription and proteotoxicity, alterations in lamin B1 expression might have important implications in HD pathophysiology.

Finally, it has been demonstrated that apoptosis is induced by inhibiting the expression of ribosomal protein S3A (RPS3A). It also has been observed that SNP variants in RPS3A homologues are associated with pathogenesis of Alzheimer’s disease. Apart from its function as a ribosomal protein, RPS3A might also act as a chaperone. Co-expression of mouse RPS3A suppressed the toxicity induced by α-synuclein (which is a major component of Lewy bodies observed in Parkinson’s disease) in a yeast model system.

As the literature review indicated, the intersection of the UPR interactome with apoptosis-related genes and HTT interactors can point out proteins with potential relevance for neurodegeneration. Thus, the generated gene lists provided in the Supplementary data file 6 and Supplementary data file 7 might give interested researchers a valuable basis for more detailed inspections.

Conclusions

Various studies have indicated a role of the UPR in HD. However, its relevance for therapeutic interventions remains to be established. With the presented work, we aimed to delineate the connection between UPR and HD by examining available HD-relevant gene expression and molecular interaction data. We found indications for differential regulation of UPR genes in a number of expression studies. Notably, the observed differential regulation is not conserved across all evaluated studies reflecting the well-known heterogeneity of current HD models. This needs to be taken into account for future studies of the UPR in the context of HD. The results of our analysis (displayed in Figure 2) may therefore serve as guidance
for the choice of model systems. Despite the observed heterogeneity, the comparison nevertheless indicated a number of genes that tend to be commonly regulated in different expression studies. This finding enabled us to define core sets of UPR genes that were commonly up- or downregulated in different studies. This derivation was supported by the detection of a significant overrepresentation of UPR-associated stress response elements (UPRE and ERSE) in the promoter regions of the upregulated genes.

Functional enrichment analysis on differentially expressed UPR genes pointed to a broad range of mechanisms involved. Additional pathway analyses indicated the activation of inflammatory processes and a potential connection to diabetes. Including complementary data sets, we identified UPR genes that have been indicated to influence HD pathogenesis. Finally, we derived sets of genes that connect UPR with apoptosis and might be directly influenced by mHTT.
In summary, through our work we present the first comprehensive analysis of UPR activation in HD and elucidate potential links to pathogenetic mechanisms within a systems biology framework. While our work cannot provide definite proofs for the identified relations due to its purely computational nature, it can nevertheless constitute a broad basis for experimental follow-up investigations. To assist such endeavours, extensive supplementary material has been provided together with this article with the aim of helping independent researchers to select genes of interest. We are also currently developing a publicly accessible web-portal for the retrieval and visualisation of changes in UPR-associated gene expression across the evaluated transcriptomics studies. In conclusion, we hope that our work can contribute to a better understanding of the UPR in HD and eventually to the identification of novel therapeutic targets to cure HD.

Data availability


Author contributions

RK collected the data, performed the analysis and prepared manuscript. JGL, SM and KA collected data and wrote parts of manuscript. MF conceived the study, contributed to interpretation of the results and wrote the final version of the manuscript. All authors agreed to the final content of the manuscript.

Competing interests

No competing interests were disclosed.

Grant information

The work presented was supported by CHDI Foundation (A-2666) and by the Portuguese Fundação para a Ciência e a Tecnologia (SFRH/BPD/70718/2010 to RKK and IF/00881/2013 to MEF).

Acknowledgements

We would like to thank Isabel Duarte for critical reading of the manuscript.

Supplementary information

**Supplementary Figure 1.** Comparison of UPR gene lists. Venn diagram comparing the three different UPR genes lists generated. **Supplementary Figure 2.** Common downregulated UPR genes. Comparison of several HD expression data sets to identify UPR genes that are downregulated in different HD models.
Supplementary Figure 3. Comparison of UPRE and ERSE elements in the promoter region of UPR genes that are upregulated in HD.

Supplementary Figure 4. UPR-HD connectors. Venn diagram showing common genes between 3 data sets, UPR-HDdiff: UPR genes that are differentially regulated in HD; HDTT: HD therapeutic targets as described by Kalathur RK et al.37 and originated by the HDCrossRoads database; and HTT-int: HTT interactors derived from the HDNetDB.

References


F1000Research 2015, 4:103 Last updated: 16 MAY 2019

Page 15 of 20


Data Source
Stefan Taubert
Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada

In this paper, Kalathur et al. investigated the role of the ER’s unfolded protein response (UPR) in the disease pathogenesis of Huntington’s Disease (HD). Although HTT, the protein mutated in HD, does not localize to the ER, mutated HTT interferes with downstream UPR processes, which may prevent the UPR from functioning normally and thus result in the induction of apoptosis by the UPR, which in turn could underlie the dramatic neuronal loss observed in HD. To broadly evaluate the role of UPR activation in the many experimental HD models and contexts, the authors used a bioinformatics approach to query for the role of the UPR in HD pathology, assessing transcriptomes, regulatory DNA elements, and UPR interactomes. The resulting data point to a core set of UPR genes that were commonly (although, as nicely pointed out, not universally) up- or downregulated in distinct experimental HD models, pro-inflammatory events involving microglia, a putative link between HD and diabetes/hyperglycemia, and the identification of a core set of genes that link UPR to apoptotic signaling and that mHTT may impinge upon.

Overall, this is a nice body of work. This computational analysis is comprehensive and multifaceted, the data interpretation measured and well-qualified, the paper well written, and the data are very accessible. One of the key strengths is that the authors aimed to integrate data obtained in many different experimental HD systems - various mouse models, human data, and even yeast and rat models. This comprehensive approach allows them to point to evolutionarily conserved genes and processes as suitable candidates for future investigation.

Specific criticism:
- To derive their set of HTT interacting proteins, the authors use a dataset downloaded from HDNetDB database, which yields a large set of HTT interactors (HTT-int) including 1015 genes, including, as acknowledged, presumably many indirect interactors. Perhaps a deeper up-front curation for putative direct interactors would have been useful.
- In Figure 2, the authors assess the overlap between genes regulated in 6 mouse HD datasets and one human. To ensure best stringency, the first limit the mouse dataset to those genes that are regulated in all 6 individual datasets. While this high stringency is laudable, one could have also
lowered the stringency a bit in regards to inclusion in the core mouse IPR set, i.e. inclusion in 6/6 sets was deemed required, but expanding the set to 5/6 or even 4/6, resulting in 420 and 902 additional genes, respectively, would have only marginally lowered the stringency while providing a larger set for the determination of evolutionary conservation.

**Competing Interests:** No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

---

**Author Response 17 Feb 2016**

Ravi Kiran Reddy Kalathur, Centre for Biomedical Research, University of Algarve, Portugal

We thank the reviewer for the positive comments and specific suggestions that helped us to further improve our manuscript. In the revised version, we have included the additional analysis suggest by the reviewer.

1. We distinguish now between direct and indirect interactions of HTT based on their annotation in HDNetDB. An additional supplementary table (Supplementary dataset 8) presents now the classification of HTT interactions. This information was also used for the interpretation of our results. For instance, 2 of the 13 proteins constituting a putative link between HD and UPR, were direct interactors. We did not want to exclude a priori all indirect HTT interactors, since those might be part of larger protein complexes with HTT and thus potentially important for the HTT’s molecular functions.

2. Additional lists of genes with evolutionary conserved expression were generated for less stringent filters (i.e. requiring the inclusion in the enrichment cores of at least 5 or 4 comparisons for up- or downregulated genes, respectively). In general, promotor and enrichment analyses of the extended gene lists agree well with those for the more stringent genes lists. The results are presented in supplementary tables and figures, and briefly discussed in the main manuscript. For the connection to HD, lowering the requirement led to detection of 53 candidates, of which 15 were direct HTT interactors (Supplementary figure 5). This enlarged set is presented in Supplementary dataset 9 for the interested reader.

**Competing Interests:** No competing interests were disclosed.
In HD and other neurodegenerative disorders involving the accumulation of misfolded protein, ER stress and the activation of UPR have been implicated in pathogenesis, although the extent to which it contributes to neuronal cell death is still unclear. In this paper, the authors employ a bioinformatic approach to show that the unfolded protein response (UPR) is activated in a variety of human and animal Huntington’s disease (HD) models. It is noteworthy that the authors also demonstrate that the upregulation of the UPR genes is most likely a direct response to UPR activation by analyzing the promoter regions of these genes for the binding sites of the UPR transcriptional activators ATF6 and XBP1. Moreover, the authors also identified sets of genes that provide a potential link between both UPR and HD and between UPR and apoptosis. Pathway enrichment analysis was also used to identify functional pathways activated by the common set of upregulated UPR genes. In addition to the expected pathways associated with UPR (e.g. PERK-regulated gene expression and ATF-activated genes), the authors also identified components of the immune system, neurotrophin signaling, and diabetes. Interestingly, there is evidence in the HD literature that all three of these latter pathways are affected in HD.

This work is a good example of the power of using a systems biology approach to provide a fairly comprehensive analysis of UPR activation in HD. The authors are appropriately cautious in emphasizing that their results do not prove that the pathways and relationships between pathways that they have identified all contribute to HD pathogenesis. However, their results provide an excellent guide for further experimental studies, and the authors’ development of a publicly accessible web site for the retrieval and visualization of their UPR-associated gene expression data in HD will be an important tool for the field that will facilitate these future studies.

Minor comments:

1. In the introduction, the authors mention that although HTT is not located in the ER, there are several potential mechanisms by which mutant HTT could induce ER stress (impairment of ERAD, dysfunctional vesicular trafficking, and altered ER calcium homeostasis). Atwal et al., (2007) and Atwal and Truant (2008) have shown that there is a more direct link between HTT and ER stress. In their work, they show that the N17 domain of HTT is a stress-sensitive ER association domain, and that the expanded polyQ stretch in mutant HTT perturbs the release of mutant HTT from ER and its translocation of HTT into and out of the nucleus in response to cell stress events.

2. p.7, 1st paragraph: “…UPR has been also associated with the suppression or degradation of a substantial number of genes…. Do the authors mean: …UPR has been also associated with the suppression or degradation of a substantial number of genes or gene products?

3. To account for the differences in differential UPR gene expression changes that were observed among the different HD mouse models, the authors suggest that differences in the HTT transgenes and their expression levels among the models could be responsible. In addition, the different genetic backgrounds of the models can contribute to the variability. In future studies aimed at studying the role of the length of mutant HTT’s expanded polyQ stretch in UPR activation, the authors may also want to consider evaluating CHDI’s publicly available transcriptome data obtained from the cortex and striatum of 6-month old knock-in HD mouse models expressing wild-type or mutant Htt alleles with different CAG repeat lengths.

**Competing Interests:** No competing interests were disclosed.
I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 17 Feb 2016

Ravi Kiran Reddy Kalathur, Centre for Biomedical Research, University of Algarve, Portugal

We would like to thank the reviewer for the encouraging evaluating of our systems biology approach and his suggestions for further improvement of the study and its presentation. In response to the comments, following changes to the manuscript were made:

1. In the Introduction, we pointed out the finding of a stress-sensitive association of HTT with the ER membrane, which suggests a potential direct role of HTT in ER stress. The relevant references are now included.

2. We clarified the meaning of the sentence. It states now: “As the UPR has also been associated with suppression of gene transcription and the enhanced degradation of numerous transcripts, ...”

3. We would like to thank reviewer for pointing out this data set. We analysed the correlation of expression changes with the length of polyglutamine tract. Notably, UPR genes were found to be strongly overrepresented among genes whose expression in the striatum significantly correlate with the length of polyglutamine tract. The correlation and the corresponding significance are also shown for the potential UPR-HD connectors in table 3 and Supplementary dataset 9. The results suggest length of the polyglutamine tract (which is reversely correlated with age of disease onset in human patients) plays a critical role in the activation of UPR. We aim to include these expression changes also in our newly developed web-portal UPRHD at [http://uprhd.sysbiolab.eu](http://uprhd.sysbiolab.eu).

Competing Interests: No competing interests were disclosed.