Human Molecular Genetics, 2018, Vol. 27, No. 17 2965–2977

doi: 10.1093/hmg/ddy198
Advance Access Publication Date: 22 May 2018
Original Article

ORIGINAL ARTICLE

Peripheral blood gene expression reveals an inflammatory transcriptomic signature in Friedreich's ataxia patients

Daniel Nachun[1], Fuying Gao[1], Charles Isaacs[2], Cassandra Strawser[2], Zhongan Yang[1], Deepika Dokuru[1], Victoria Van Berlo[1], Renee Sears[1], Jennifer Farmer[3], Susan Perlman[4], David R. Lynch[2] and Giovanni Coppola[1,4,*]

[1]Department of Psychiatry and Semel Institute, University of California, Los Angeles, Los Angeles, CA 90095, USA, [2]Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA, [3]Friedreich's Ataxia Research Alliance, Downingtown, PA 19104, USA and [4]Department of Neurology, University of California, Los Angeles, Los Angeles, CA 90095, USA

*To whom correspondence should be addressed. Tel: 310-794-4172; Fax: 310-794-9613, Email: gcoppola@ucla.edu

Abstract

Transcriptional changes in Friedreich's ataxia (FRDA), a rare and debilitating recessive Mendelian neurodegenerative disorder, have been studied in affected but inaccessible tissues—such as dorsal root ganglia, sensory neurons and cerebellum—in animal models or small patient series. However, transcriptional changes induced by FRDA in peripheral blood, a readily accessible tissue, have not been characterized in a large sample. We used differential expression, association with disability stage, network analysis and enrichment analysis to characterize the peripheral blood transcriptome and identify genes that were differentially expressed in FRDA patients ($n = 418$) compared with both heterozygous expansion carriers ($n = 228$) and controls ($n = 93$ 739 individuals in total), or were associated with disease progression, resulting in a disease signature for FRDA. We identified a transcriptional signature strongly enriched for an inflammatory innate immune response. Future studies should seek to further characterize the role of peripheral inflammation in FRDA pathology and determine its relevance to overall disease progression.

Introduction

Friedreich's ataxia (FRDA, OMIM 229300) is a rare autosomal recessive disorder characterized by progressive ataxia, significant loss of motor control, cardiomyopathy and diabetes. The disorder is usually caused by an intronic trinucleotide (GAA) repeat expansion in the highly conserved gene frataxin (*FXN*, ENSG00000165060), whose protein product is essential to the formation of iron-sulfur cluster complexes. These complexes are necessary for the proper functioning of a large number of proteins, particularly those involved in mitochondrial metabolism. FRDA is a result of *FXN* haploinsufficiency, and complete loss of *FXN* is embryonic lethal (1). FRDA patients exhibit a 70–80% reduction of *FXN* expression levels compared with unaffected individuals (2). Heterozygous expansion carriers exhibit a

Received: October 18, 2017. **Revised:** March 30, 2018. **Accepted:** May 17, 2018

modest reduction in *FXN* expression and do not develop clinical symptoms.

FXN deficiency causes a number of pathologies at the cellular level (reviewed in 3). A large build up in mitochondrial iron and reduced function of antioxidant proteins lead to an increase in reactive oxygen species, which lead to severe oxidative stress characterized by damage to proteins, DNA and lipid membranes. These effects can ultimately lead to degeneration and cell death, particularly in post-mitotic cells with very high metabolic activity, such as large neurons, cardiomyocytes and pancreatic islet cells (4), but many of the affected pathways are universal to the function of all eukaryotic cells, and a more subtle transcriptional response may be present in peripheral tissues not clinically involved, but readily available for study in large cohorts. In addition, because FRDA results in severe metabolic stress and eventual loss of cells of the peripheral and central nervous system, this may lead to a peripheral immune response that can be detected at the level of gene expression in blood immune cells. To explore these hypotheses, we collected the largest series to date of RNA from peripheral blood from FRDA patients, carriers, and controls, and performed microarray-based gene expression analysis. We identified an inflammatory disease-associated signature which in part overlaps with previous datasets from patients and animal models. The entire dataset is available to the FRDA community in a web-based application (REPAIR) for data mining and additional analyses.

Subjects and Samples

A total of 739 subjects were enrolled at two sites, UCLA and the Children's Hospital of Philadelphia (CHOP). Table 1 provides a basic summary of the demographics of the subjects. Subjects were divided into three groups based on clinical diagnosis. Patients were those subjects clinically diagnosed with FRDA ($n = 418$) and in most (90.6%) the approximate number of GAA repeats in the *FXN* gene was also determined via PCR (5) to serve as molecular confirmation, as well as an indirect measure of disease severity. Eighteen patients were compound heterozygotes with one repeat expansion and one loss-of-function point mutation in *FXN* (Supplementary Material, Table S1). Carriers were those subjects carrying one expanded *FXN* allele and one normal allele ($n = 228$). Most carriers were parents of patients, who are obligate carriers. Control subjects consisted of individuals known not to have any relatives with FRDA. Because carriers and controls are phenotypically indistinguishable, we checked blood frataxin levels in 95 enrolled controls (6) and excluded 2 subjects with frataxin levels lower than the range observed in homozygote expansion carriers, leaving 93 controls for further analyses.

Results

Differential expression

In order to identify a peripheral signature related to FRDA pathology, we fit linear models for each transcript. At a cut-off of $\log_{10}$ Bayes factor > 0.5 (log BF, see Materials and Methods) comparing the alternative model containing disease status to the null model without it, after accounting for a number of potential confounders (see Materials and Methods and Supplementary Material, Figs S1 and S2), 1115 transcripts were significant for the effect of disease status across all three groups. To identify transcripts that were significantly differentially expressed (DE) across pairwise comparisons, we computed posterior probabilities and identified transcripts for each pairwise comparison where the posterior probability (pp) of differential expression was greater than 0.95 (see Materials and Methods). The global false discovery rate (FDR) for each set of DE transcripts in each comparison was also computed as described in Materials and Methods. Of the 1115 transcripts identified as being significantly affected by disease status, 829 transcripts were DE between patients and controls (global FDR = 0.012), 1078 between patients and carriers (global FDR = 0.0017) and 182 between carriers and controls (global FDR = 0.018) (Fig. 1, Table 2, Supplementary Material, Table S2). The observation that more genes were DE in patients versus carriers compared with patients versus controls is likely due to the much larger number of carriers (228) compared with controls (93), which provides stronger statistical support to small expression changes.

Table 1. Summary of subject demographics

Status	Male	Female	Total	Age	GAA1 length
Patient	221 (53%)	197 (47%)	418	25 ± 11.9	900 ± 185
Carrier	89 (39%)	139 (61%)	228	50 ± 17.8	N/A
Control	53 (57%)	40 (43%)	93	37 ± 10.4	N/A
Total	363	376	739		

Regression with clinical phenotypes

Several phenotypic measures can be used to quantify disease severity in FRDA patients. A direct clinical measure is the functional disability stage (FDS) score developed for the Friedreich's ataxia rating scale (FARS) (7), which rates patients on a scale from 0–6 based upon their mobility, with 0 indicating no impairment and 6 complete disability. Two less direct measures of disease severity are the disease duration in years and the size of the shorter GAA repeat expansion in patients, GAA1. We used linear modeling to identify transcripts with significant positive or negative linear relationships with each phenotypic measure. At a cut-off of log BF > 0.5, comparing the alternative model with the phenotypic measure to the null model without it, we identified 1508 transcripts significantly associated with FDS (global FDR = 0.0028, Fig. 2, Table 3, Supplementary Material, Table S3), 280 transcripts significantly associated with GAA1 (global FDR = 0.0043) and 13 transcripts significantly associated with disease duration (global FDR = 0.006). In all three analyses, all genes with log BF > 0.5 also had a pp > 0.95.

Enrichment analysis of DE and FDS-associated genes

We used enrichment analysis to identify biological pathways that were significantly overrepresented in DE or FDS-associated genes (Fig. 3). In genes that were significantly upregulated in patients compared with carriers and controls, we identified a very strong enrichment for one specific process: neutrophil degranulation (patient vs. control: 58 genes, log BF = 22.2, patient vs. carrier: 70 genes, log BF = 26.6). There was weaker enrichment for downregulated genes in general, with the strongest term relating to T-cell differentiation (patient vs. control: 12 genes, log BF = 6.26, patient vs. carrier: 14 genes, log BF = 6.67). This enrichment is supported by the presence of numerous T-cell marker genes (*CCR7*, *CD8A*, *GZMK*, *CD3D*, *CD27*) in the most downregulated genes in patients. These results indicate the presence of subtle but robust changes in peripheral blood gene expression associated with the presence of a pathogenic mutation in *FXN*.

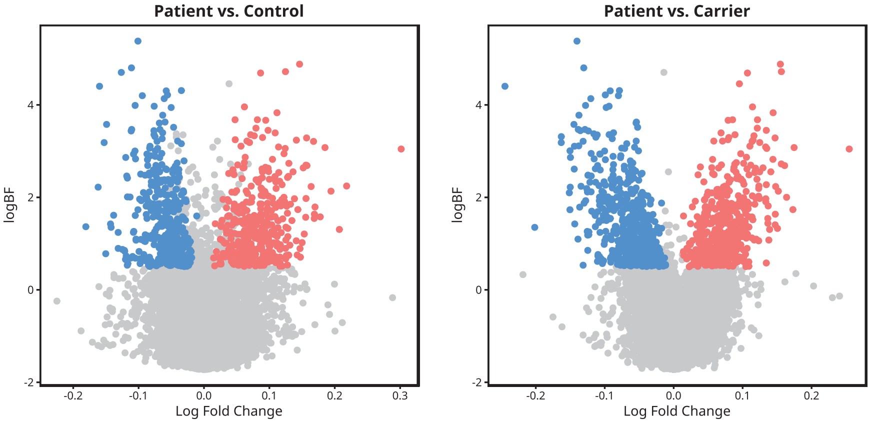


Figure 1. Differential expression analysis identifies 829 genes DE between patients and controls and 1078 genes DE between patients and carriers. Volcano plot of all genes in patient versus control and patient versus carrier comparisons. The fold change is on the *x*-axis, and the log BF is on the *y*-axis. Blue indicates a gene that is significantly (log BF > 0.5, pp > 0.95) downregulated, while red indicates a gene that is significantly upregulated.

Table 2. Top genes most DE between patients and controls

Gene	Definition	Log BF	Log FC (pat. vs. cont.)	Log FC (pat. vs. carr.)	Function
MMP9	Matrix metallopeptidase 9 (ENSG00000100985)	3.05	0.302	0.255	(UniProt KB) May play an essential role in local proteolysis of the extracellular matrix and in leukocyte migration.
ANPEP	Alanyl aminopeptidase, membrane (ENSG00000166825)	2.25	0.218	0.124	(UniProt KB) Broad specificity aminopeptidase.
DYSF	Dysferlin (ENSG00000135636)	1.31	0.207	0.139	(UniProt KB) Key calcium ion sensor involved in the Ca(2+)-triggered synaptic vesicle-plasma membrane fusion.
MME	Membrane metalloendopeptidase (ENSG00000196549)	2.13	0.194	0.152	(Entrez Gene) The protein is a neutral endopeptidase that cleaves peptides at the amino side of hydrophobic residues.
RPL14	Ribosomal protein L14 (ENSG00000188846)	1.36	−0.181	−0.0541	Ribosome component.
RNF24	Ring finger protein 24 (ENSG00000101236)	1.58	0.178	0.0965	(Entrez Gene) This gene encodes an integral membrane protein that contains a RING-type zinc finger.
PADI4	Peptidyl arginine deiminase 4 (ENSG00000159339)	3.08	0.185	0.175	(Entrez Gene) This gene is a member of a gene family which encodes enzymes responsible for the conversion of arginine residues to citrulline residues.
NCF4	Neutrophil cytosolic factor 4 (ENSG00000100365)	1.79	0.170	0.107	(UniProt KB) Component of the NADPH-oxidase, a multicomponent enzyme system responsible for the oxidative burst.
CA4	Carbonic anhydrase 4 (ENSG00000167434)	1.54	0.169	0.137	(UniProt KB) Reversible hydration of carbon dioxide.
LRRN3	Leucine-rich repeat neuronal 3 (ENSG00000173114)	3.18	−0.153	−0.163	Unannotated.

Annotations provided by GeneCards (RRID: SCR_002773), UniProt (RRID: SCR_002380) and Entrez Gene (RRID: SCR_002473). Log FC = $\log_2$ fold change, log BF = $\log_{10}$ Bayes factor.

Remarkably, enrichment analysis identified the same top term for genes positively associated with FDS: neutrophil degranulation (38 genes, log BF = 5.78). Negatively associated genes had weaker overall enrichment, which was primarily centered around RNA processing (mRNA splicing: 39 genes, log BF = 4.28; tRNA modification: 12 genes, log BF = 3.8; rRNA modification: 32 genes, log BF = 3.25). Although not significantly enriched, several of the most negatively associated genes (*CD79A*, *GZMB*) are also lymphocyte marker genes, recapitulating the decrease in similar genes seen in differential expression.

Overlap with other datasets

Gene expression changes associated with frataxin deficiency have previously been studied in a number of models, including transgenic mice, as well as peripheral blood. Two human datasets from peripheral blood (GSE11204, GSE30933, see Supplementary Material for descriptions of each dataset), and one mouse dataset (RNAi mouse) (8) were analyzed using the same differential expression workflow used with our data. We considered upregulated (log FC > 0) and downregulated (log FC < 0) transcripts separately (or positive and negative

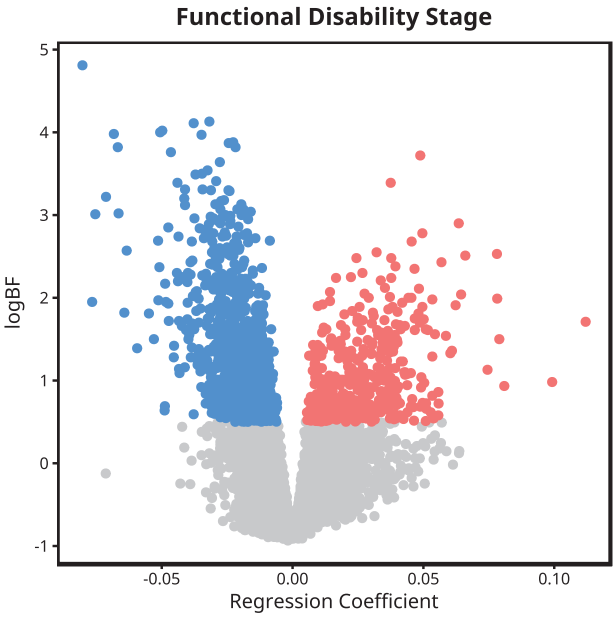


Figure 2. Regression of gene expression with FDS identifies 1508 transcripts significantly associated with FDS. Volcano plot of all genes in FDS regression. The regression coefficient is on the x-axis, and the log BF is on the y-axis. Blue indicates a gene with a significant (log BF > 0.5) negative regression coefficient, while red indicates a gene with a significant positive coefficient.

regression coefficients for FDS-associated transcripts) and the overlaps were computed for patients versus controls, patients versus carriers and FDS regression in each direction of change. Thirteen comparisons had a log BF greater than 0.5 (Fig. 4), indicating that our DE and FDS-associated genes were significantly enriched for genes enriched in differential expression in other datasets.

In all cases, the overlap was only observed in the upregulated genes. Six of the enriched comparisons originate from the patient versus control and carrier versus control contrasts from a previously published peripheral blood dataset (GSE30933), while the other seven enrichments are seen in DE genes in heart tissue collected at several developmental timepoints in a novel mouse model of FRDA (8). No enrichment was observed for the other previously published peripheral blood dataset (GSE11204). There was also no enrichment observed in DE genes in cerebellum and dorsal root ganglion (DRG) tissue collected from the same mouse model (Supplementary Material, Fig. S3).

Weighted gene coexpression network analysis (WGCNA)

WGCNA is a powerful method for the identification of groups of coexpressed transcripts (9–12). We first identified modules in our dataset, then used module eigengenes (see Materials and Methods) as summary measures for each module to determine if any modules were significantly different across our diagnostic groups, or related to disease progression, using the same linear model designs as in the previous analyses.

Diagnosis

First, we assessed the relationship with diagnostic groups. We identified seven distinct coexpression modules in the complete dataset (Fig. 5A and B, Supplementary Material, Table S4). Three of the seven modules had a log BF > 0.5 for the alternative model compared with the null (Fig. 5C): pink (log BF = 2.17), green (log BF = 1.38) and black (log BF = 1.67). To identify the specific pairwise differences in the eigengene values, we also computed posterior probabilities and FDRs for the contrasts previously described for differential expression (patient vs. control FDR = 0.0015, patient vs. carrier FDR = 0.0017). The pink module eigengene was significantly higher in patients than in controls (log FC = 0.011, pp = 0.994) and carriers (log FC = 0.012, pp = 1.0), while no difference was observed between carriers and controls. The green module eigengene was also higher in patients compared with controls (log FC = 0.012, pp = 0.998) and carriers (log FC = 0.009, pp = 1.0). Conversely, the black module eigengene was significantly decreased in patients compared to controls (log FC = −0.008, pp = 0.970) and carriers (log FC = −0.012, pp = 1.0). The top hub genes in each module are listed in Table 4.

Functional disability stage

We also used WGCNA to identify groups of coexpressed genes correlated with FDS. Using the same subset of patients as in the regression with FDS, we identified eight modules (Fig. 6A and B, Supplementary Material, Table S5), and used the same linear model designs described for regression with FDS to determine if any eigengenes were significantly associated with FDS. Three modules had a log BF > 0.5 for the alternative model compared with the null (global FDR = 1.0 × 10^{-4}, Fig. 6C): the magenta module (coeff. = 0.0071, log BF = 1.45, pp = 0.999), the yellow module (coeff. = −0.0093, log BF = 3.01, pp = 1.0) and the red module (coeff. = −0.0077, log BF = 1.91, pp = 1.0). The top hub genes for each module are listed in Table 5. Remarkably, the top three hub genes of the magenta module are the same as the top three hub genes of the pink module from the status network, and two of the top three hub genes in the yellow module are shared with the black module in the status network.

Enrichment analysis of significant modules

Similar to the approach taken with DE and FDS-associated genes, we used enrichment analysis to identify biological pathways which were overrepresented in our significant WGCNA modules (Fig. 7). In the status network, the pink module was highly enriched for neutrophil degranulation (43 genes, log BF = 12.0), the same process seen in upregulated genes in differential expression and genes positively associated with FDS. The green module exhibited even stronger enrichment for neutrophil degranulation (156 genes, log BF = 38.8). The likely reason the green module is separate from the pink module is that the green eigengene is slightly increased in carriers, while the pink module shows no difference between carriers and controls. Finally, the black module, while showing weaker enrichment overall, did contain a large number of genes involved in rRNA modification (49 genes, log BF = 1.47).

In the FDS network, we found that the magenta module was strongly enriched for the same inflammatory response, neutrophil degranulation (75 genes, log BF = 14.2), as seen in the pink module in the status network. Enrichment analysis of yellow module indicated enrichment for rRNA processing (44 genes, log BF = 2.98) and the mitochondrial respiratory chain complex (20 genes, log BF = 1.89). Finally, the red module was strongly enriched for translation, especially mitochondrial translation (27 genes, log BF = 4.44).

Cell type deconvolution

Changes in cell type composition could in theory explain some of the changes in gene expression we observed between

Table 3. Top genes most strongly associated with FDS

Gene	Definition	Log BF	FDS coefficient	Function
PI3	Peptidase inhibitor 3 (ENSG00000124102)	1.71	0.11	(UniProt KB) This gene encodes an elastase-specific inhibitor that functions as an antimicrobial peptide.
CA1	Carbonic anhydrase 1 (ENSG00000133742)	0.98	0.10	(UniProt KB) Reversible hydration of carbon dioxide. Can hydrates cyanamide to urea.
SNCA	Synuclein alpha (ENSG00000145335)	0.93	0.081	(UniProt KB) Reduces neuronal responsiveness to various apoptotic stimuli, leading to a decreased caspase-3 activation.
FCGBP	Fc fragment of IgG-binding protein (ENSG00000275395)	4.81	−0.080	(UniProt KB) May be involved in the maintenance of the mucosal structure as a gel-like component of the mucosa.
GNG10	G Protein Subunit Gamma 10 (ENSG00000242616)	1.50	0.079	(UniProt KB) Guanine nucleotide-binding proteins (G proteins) are involved as a modulator or transducer in various transmembrane signaling systems.
PROK2	Prokineticin 2 (ENSG00000163421)	1.99	0.078	(UniProt KB) May function as an output molecule from the suprachiasmatic nucleus (SCN) that transmits behavioral circadian rhythm.
CHPT1	Choline phosphotransferase 1 (ENSG00000111666)	2.53	0.078	(UniProt KB) Catalyzes phosphatidylcholine biosynthesis from CDP-choline.
GZMB	Granzyme B (ENSG00000100453)	1.95	−0.077	(UniProt KB) This enzyme is necessary for target cell lysis in cell-mediated immune responses.
CD79A	CD79a molecule (ENSG00000105369)	3.01	−0.075	(EntrezGene) This gene encodes the Ig-alpha protein of the B-cell antigen component.
ALPL	Alkaline phosphatase, liver/bone/kidney (ENSG00000162551)	1.13	0.074	(EntrezGene) This gene encodes a member of the alkaline phosphatase family of proteins.

Annotations provided by GeneCards (RRID: SCR_002773), UniProt (RRID: SCR_002380) and Entrez Gene (RRID: SCR_002473). The regression coefficients are not directly comparable with correlation coefficients because the expression values and FDS cannot be standardized in a linear model including covariates. However, the relative magnitude of the coefficient still reflects the strength of the linear relationship. Log BF = $\log_{10}$ Bayes factor.

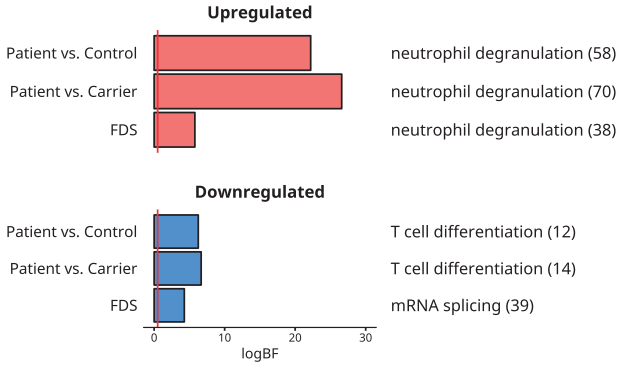


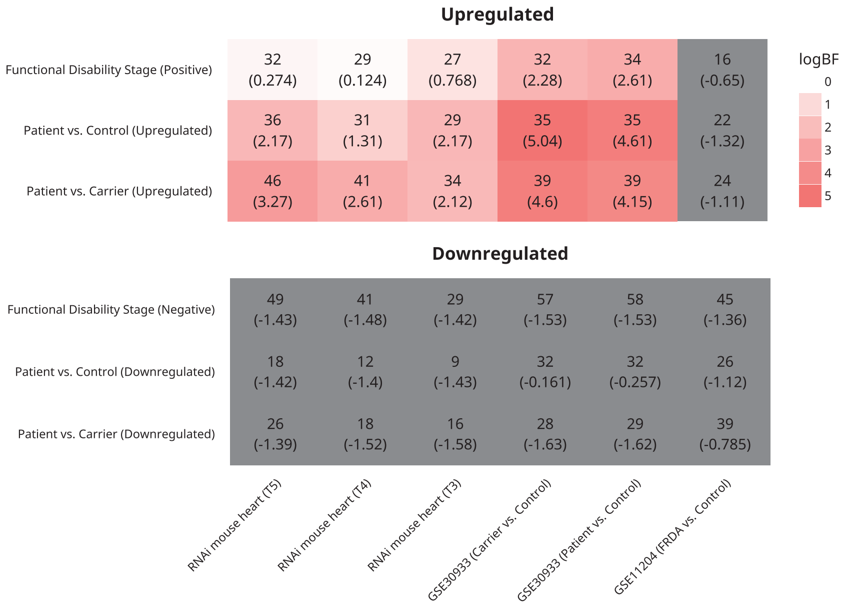
Figure 3. Enrichment analysis identifies biological pathways that are significantly overrepresented in DE and FDS-associated genes. Bar plot of most representative enrichment term for each gene set on the y-axis. The label on the right is the pathway, and the number in parentheses is the size of the overlap between the gene set and pathway. The log BF is on the x-axis, and is statistically significant at log BF > 0.5 (marked by red line).

patients, carriers, and controls and with FDS. The availability of cell-type specific transcriptomes in well-studied tissues such as peripheral blood has led to the development of tools to estimate the proportion of cell types in a sample known to contain a mixed population of cells. We used the *CellMix* tool (13) with an existing cell-type specific peripheral blood dataset (14) to estimate cell type proportions in our full dataset (patients, carriers and controls) and the subset of FRDA patients we used for regression of gene expression with FDS, after regressing out the effects of collinear variables.

After comparing cell type proportions in our three disease status groups, only the proportion of natural killer cells was significantly different (log BF = 2.45, Fig. 8) and pairwise testing found the proportion underwent a small but significant decrease in patients compared with both carriers and controls (patient vs. control: diff. = −0.0159, pp = 1.0; patient vs. carrier: diff. = −0.0094, pp = 0.999). We also regressed cell type proportion with FDS but found no significant associations (Supplementary Material, Fig. S4).

qPCR and array validation

We validated the differential expression changes of the top three DE genes between patients and controls, *MMP9*, *DYSF* and *ANPEP*, using quantitative polymerase chain reaction (qPCR) in 32 patients (including 21 additional samples not previously included in the analysis) and 32 age and sex-matched controls (including 16 new samples, Fig. 9). We also analyzed the corresponding array data for samples for which this was available (14/32 controls and 22/32 patients). In the qPCR data, there were no significant differences between patients and controls for *MMP9* ($P < 0.08$, log FC = −0.0006, Mann–Whitney *U*-test), *DYSF* ($P < 0.76$, log FC = 0.022) or *ANPEP* ($P < 0.26$, log FC = −0.007). In the corresponding array data, *MMP9* was significantly increased in patients ($P < 0.013$, log FC = 0.92), while *ANPEP* ($P < 0.13$, log FC = 0.41) and *DYSF* ($P < 0.34$, log FC = 0.15) were upregulated but did not reach statistical significance. These results show that we have biologically validated our results with a small number of independent microarrays, but that qPCR is less powered to



Upregulated

	RNAi mouse heart (T5)	RNAi mouse heart (T4)	RNAi mouse heart (T3)	GSE30933 (Carrier vs. Control)	GSE30933 (Patient vs. Control)	GSE11204 (FRDA vs. Control)
Functional Disability Stage (Positive)	32 (0.274)	29 (0.124)	27 (0.768)	32 (2.28)	34 (2.61)	16 (-0.65)
Patient vs. Control (Upregulated)	36 (2.17)	31 (1.31)	29 (2.17)	35 (5.04)	35 (4.61)	22 (-1.32)
Patient vs. Carrier (Upregulated)	46 (3.27)	41 (2.61)	34 (2.12)	39 (4.6)	39 (4.15)	24 (-1.11)

Downregulated

	RNAi mouse heart (T5)	RNAi mouse heart (T4)	RNAi mouse heart (T3)	GSE30933 (Carrier vs. Control)	GSE30933 (Patient vs. Control)	GSE11204 (FRDA vs. Control)
Functional Disability Stage (Negative)	49 (-1.43)	41 (-1.48)	29 (-1.42)	57 (-1.53)	58 (-1.53)	45 (-1.36)
Patient vs. Control (Downregulated)	18 (-1.42)	12 (-1.4)	9 (-1.43)	32 (-0.161)	32 (-0.257)	26 (-1.12)
Patient vs. Carrier (Downregulated)	26 (-1.39)	18 (-1.52)	16 (-1.58)	28 (-1.63)	29 (-1.62)	39 (-0.785)

Figure 4. Overlap of DE genes with other datasets. The number in the top of each cell in the heatmap is the number of transcripts in the overlap and the number in parentheses is the log BF of a hypergeometric overlap test. Log BF > 0.5 is considered significant. T3 = 12 weeks old, T4 = 16 weeks old, T5 = 20 weeks old. See Supplementary Material for additional descriptions of the datasets and analytic procedures.

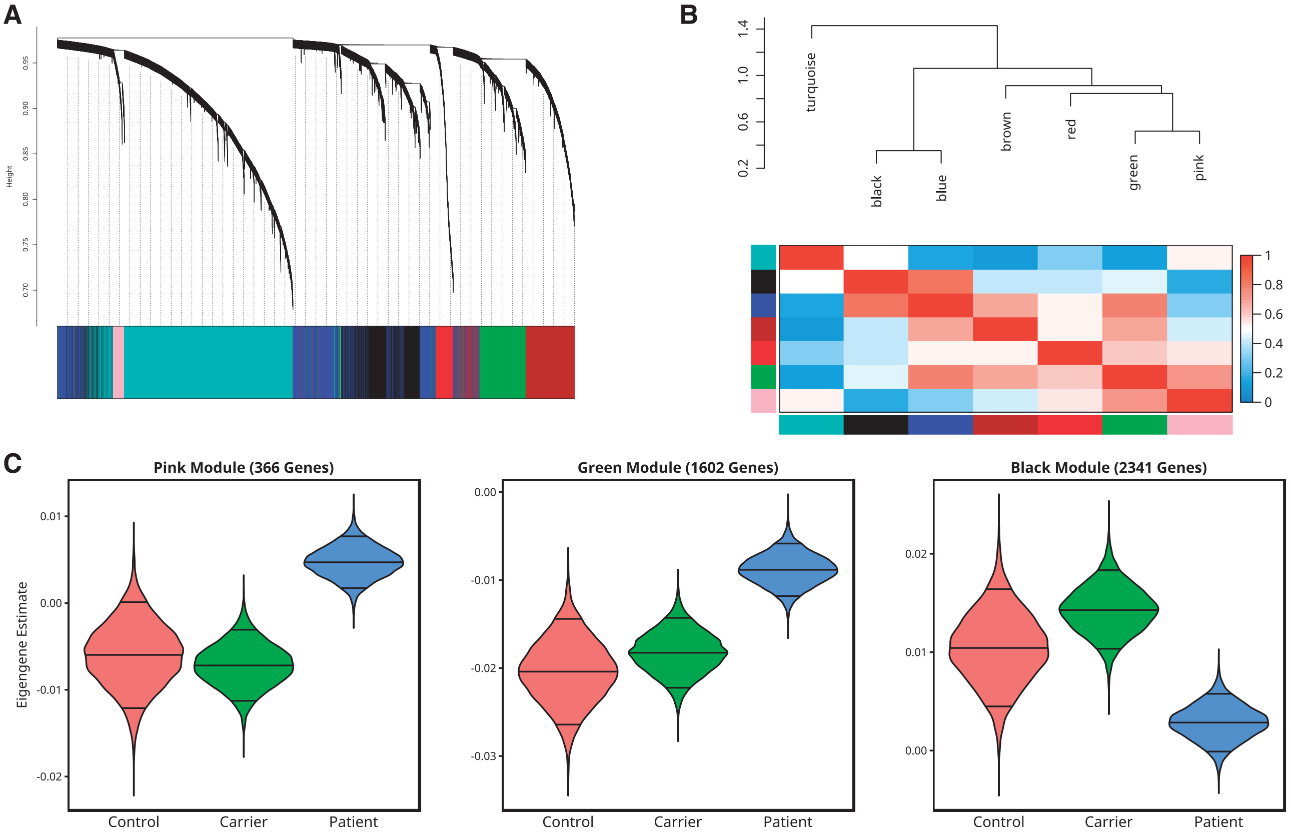


Figure 5. WGCNA identifies the pink, green and black modules as significantly different across clinical status. (**A**) Cluster dendrogram and color assignment for all transcripts in the full dataset. (**B**) Cluster dendrogram and heatmap of eigengene correlations. (**C**) Violin plots showing eigengene posterior estimates for the pink, green and black modules. The 95% credible intervals are between the smaller top and bottom lines and median estimate is the larger middle line.

Table 4. Top hub genes for the pink, green and black modules

Gene	Definition	Module	kME	Function
STX3	Syntaxin 3 (ENSG00000166900)	Pink	0.91	(UniProt KB) Potentially involved in docking of synaptic vesicles at presynaptic active zones.
MXD1	MAX dimerization protein 1 (ENSG00000059728)	Pink	0.90	(Entrez Gene) This gene encodes a member of the Myc superfamily of basic helix-loop-helix leucine zipper transcriptional regulators.
AQP9	Aquaporin 9 (ENSG00000103569)	Pink	0.89	(Entrez Gene) The aquaporins are a family of water-selective membrane channels.
DYSF	Dysferlin (ENSG00000135636)	Green	0.88	(UniProt KB) Key calcium ion sensor involved in the Ca(2+)-triggered synaptic vesicle-plasma membrane fusion.
ARID3A	AT-rich interaction domain 3A (ENSG00000116017)	Green	0.88	(UniProt KB) Transcription factor which may be involved in the control of cell cycle progression
MBOAT7	Membrane-bound O-acyltransferase domain containing 7 (ENSG00000125505)	Green	0.88	(UniProt KB) Acyltransferase which mediates the conversion of lysophosphatidylinositol into phosphatidylinositol.
FBXO31	F-box protein 31 (ENSG00000103264)	Black	0.88	(UniProt KB) Component of some SCF (SKP1-cullin-F-box) protein ligase complex that plays a central role in G1 arrest following DNA damage.
MMS19	MMS19 homolog, cytosolic iron-sulfur assembly component(ENSG00000155229)	Black	0.85	(UniProt KB) Key component of the cytosolic iron-sulfur protein assembly (CIA) complex.
USP5	Ubiquitin-specific peptidase 5 (ENSG00000111667)	Black	0.84	(UniProt KB) Cleaves linear and branched multiubiquitin polymers with a marked preference for branched polymers.

Annotations provided by GeneCards (RRID: SCR_002773), UniProt (RRID: SCR_002380) and Entrez Gene (RRID: SCR_002473). kME = correlation of gene with module hub gene.

detect small expression differences between patients and controls, likely because of small sample size and noisier quantification.

Discussion

We report the first large-scale analysis of peripheral gene expression in patients with FRDA, heterozygous mutation carriers and controls. After conservative data processing and strict statistical thresholds, we identified the transcripts with either robust differences between patients and controls, or correlated with FDS (Tables 2 and 3). In addition, network methods identified coordinated groups of genes with biological significance.

The most striking finding across our analyses was the robust enrichment for increased expression in patients of inflammatory genes, particularly those involved in neutrophil degranulation, an important innate response to tissue injury and infection which has also been implicated in chronic inflammation (15). It is not possible to determine from this data whether the inflammatory response observed peripherally is part of the disease pathogenesis, or merely a response to stress induced by *FXN* deficiency. In other chronic inflammatory disorders, activation of neutrophils and other components of the innate immune response is a key component of the disease (16). A growing body of literature also supports the involvement of both innate and adaptive immune responses in neurodegeneration, including Parkinson's disease, Alzheimer's disease and frontotemporal dementia (17). Many of our top DE genes and network hub genes are clearly linked to the innate immune response. These include several peptidases (*MMP9*, *ANPEP*, *MME*), a regulator of peptidase activity (*PI3*), two carbonic anhydrases (*CA1*, *CA4*) and genes regulating neutrophil degranulation (*NCF4*, *DYSF*, *STX3*).

We also identified a strong enrichment for a decrease in transcription and translation associated with FRDA, both when comparing patients to controls and carriers, and when examining the relationship with disease severity. Our DE genes and hub genes include *RPL14*, a ribosome component, as well as a chaperone protein (*TTC4*), and an rRNA processing gene (*DDX47*). It has long been known that oxidative stress, like that induced by FXN deficiency, leads to a decrease in translation (18), which may explain these changes.

Several of our DE or hub genes have been identified as being relevant to other neurodegenerative disorders. Both *PROK2* and *AQP9* were identified as being DE in peripheral blood in Huntington's disease (19). Mutations in dysferlin (*DYSF*) have been identified as a cause of limb-girdle muscular dystrophy (20), and mutations in alpha-synuclein (*SNCA*) have been identified in familial cases of Parkinson's disease (21). However, the relevance of these genes to the pathology of FRDA cannot be ascertained from this study.

We observed a fairly consistent overlap with our previous independent peripheral blood study including 41 subjects. The GSE11204 dataset, while also partly collected in peripheral blood (the part of the dataset collected from cell lines was not analyzed because phenotypic data were not available), was severely confounded by batch effect which might explain the poor overlap. The intriguing overlap with genes that are DE in the heart of a novel mouse model for FRDA may indicate there are some similar inflammatory processes occurring in the heart. We speculate that the complete lack of overlap with corresponding CNS tissues (DRG and cerebellum) in the same FRDA mouse model is caused by large differences in structure and function between cells of the CNS and peripheral blood and the smaller number of genes identified as DE in CNS tissues of the mouse model compared to the heart.

The detection of large numbers of genes significantly associated with FDS is intriguing given that this is a high-level clinical measurement and was collinear with age (whose effects were removed from the data before regressing with FDS). Although it is less sensitive than FARS, FDS is easier to collect in large series, and is a fairly direct measurement of disease severity, so it is biologically plausible that genes would be positively or

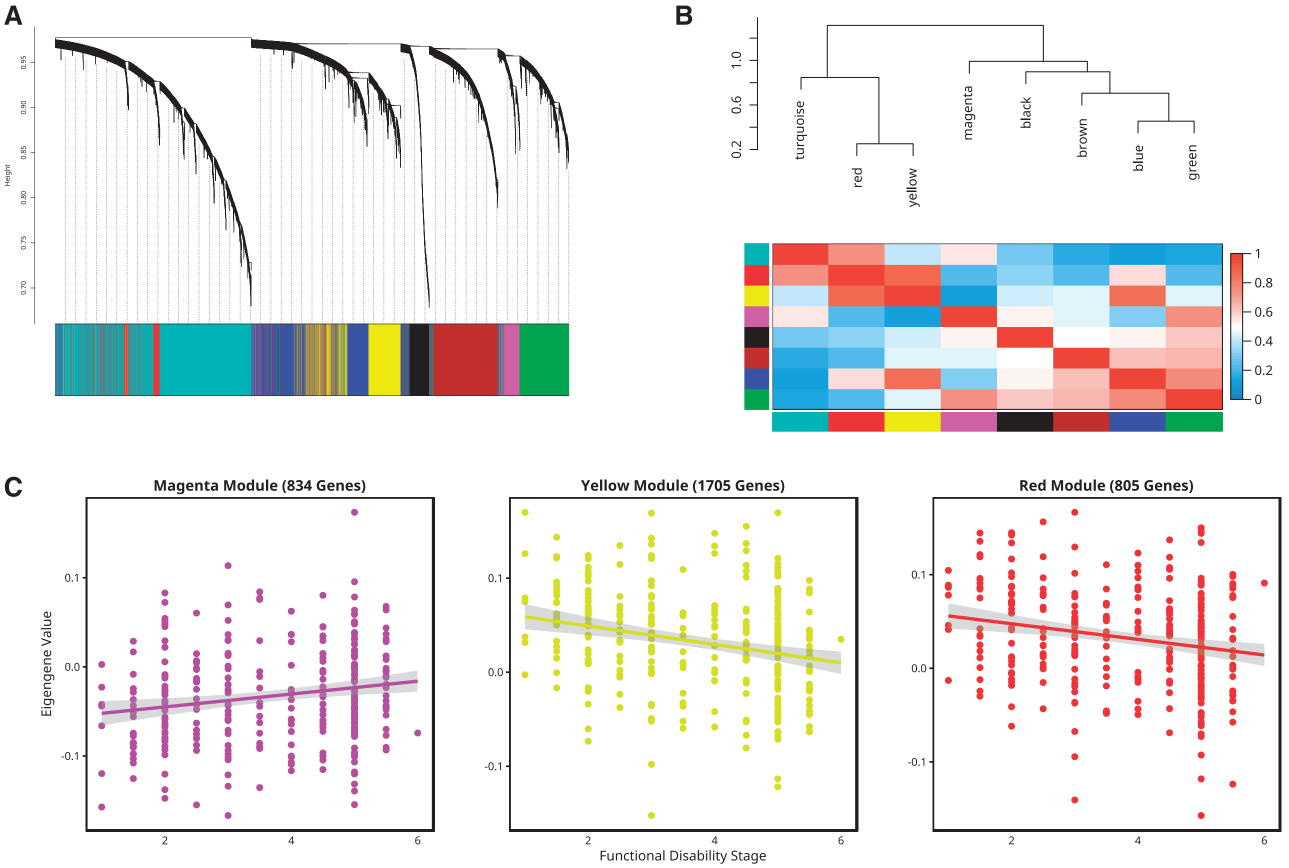


Figure 6. WGCNA identifies the magenta, yellow and red modules as significantly associated with FDS (**A**) Cluster dendrogram and color assignment for all transcripts in the patients with FDS available. (**B**) Cluster dendrogram and heatmap of eigengene correlations. (**C**) Scatterplots showing relationship of FDS (on the x-axis) with eigengene expression (on the y-axis) for the magenta, yellow, and red modules.

Table 5. Top hub genes for the magenta, red, and yellow modules

Gene	Definition	Module	Kme	Function
STX3	Syntaxin 3 (ENSG00000166900)	Magenta	0.90	(UniProt KB) Potentially involved in docking of synaptic vesicles at presynaptic active zones.
MXD1	MAX dimerization protein 1 (ENSG00000059728)	Magenta	0.91	(Entrez Gene) This gene encodes a member of the Myc superfamily of basic helix-loop-helix leucine zipper transcriptional regulators.
AQP9	Aquaporin 9 (ENSG00000103569)	Magenta	0.90	(Entrez Gene) The aquaporins are a family of water-selective membrane channels.
TTC4	Tetratricopeptide repeat domain 4 (ENSG00000243725)	Red	0.87	(Entrez Gene) This gene encodes a protein that contains tetratricopeptide (TPR) repeats, which often mediate protein–protein interactions and chaperone activity.
PPP3CC	Protein phosphatase 3 catalytic subunit gamma (ENSG00000120910)	Red	0.84	(UniProt KB) Calcium-dependent, calmodulin-stimulated protein phosphatase.
DDX47	DEAD-box helicase 47 (ENSG00000213782)	Red	0.84	(UniProt KB) Involved in apoptosis. May have a role in rRNA processing and mRNA splicing.
FBXO31	F-box protein 31 (ENSG00000103264)	Yellow	0.88	(UniProt KB) Component of some SCF (SKP1-cullin-F-box) protein ligase complex that plays a central role in G1 arrest following DNA damage.
USP5	Ubiquitin-specific peptidase 5 (ENSG00000111667)	Yellow	0.86	(UniProt KB) Cleaves are linear and branched multiubiquitin polymers with a marked preference for branched polymers.
RPUSD2	RNA pseudouridylate synthase domain containing 2 (ENSG00000166133)	Yellow	0.86	Unannotated.

Annotations provided by GeneCards (RRID: SCR_002773), UniProt (RRID: SCR_002380) and Entrez Gene (RRID: SCR_002473). kME = correlation of gene with module hub gene.

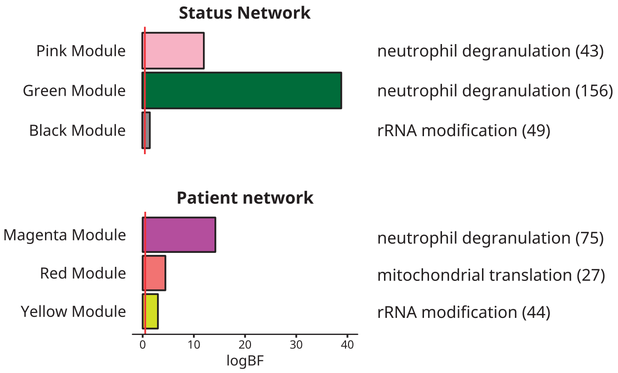


Figure 7. Enrichment analysis identifies biological pathways that are significantly overrepresented in WGCNA modules. Bar plot of most representative enrichment term for each gene set is on the y-axis. The label on the right is the pathway, and the number in parentheses is the size of the overlap between the gene set and pathway. The log BF on the x-axis, and is statistically significant at log BF > 0.5.

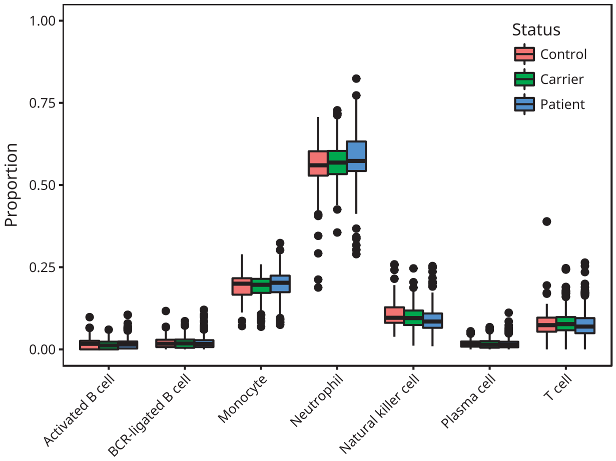


Figure 8. Cell type deconvolution analysis. Boxplots showing cell type proportion of seven cell types in patients, carriers and controls.

negatively associated with it. We were also intrigued to note that the same inflammatory response which appears to differentiate patients from controls and carriers is also positively associated with disease severity. By contrast, the enrichments seen for downregulated genes in patients and genes inversely associated with severity were generally weaker, though still consistently including transcription.

The relatively poor detection of genes associated with GAA1 or disease duration is likely to due to several issues. Both measures were collinear with age and are not direct measures of disease severity. Furthermore, somatic mosaicism may introduce differences in GAA1 length in blood compared with affected tissues such as the spinal cord, heart or pancreas.

Although we found enrichment for a number of cell type-specific signatures in our data, cell type deconvolution revealed no change in proportion of neutrophils as estimated from gene expression data, leading us to hypothesize that the large increase in neutrophil degranulation persistently seen across different analyses is not due to an absolute change in neutrophil count. We observed only a small decrease in natural killer cells in patients, which may explain the decrease in lymphocyte activation observed in differential expression, although alterations in adaptive immune responses have been observed in neurodegenerative disease (17). Complete blood cell counts should be used to properly characterize what changes, if any, occur in cell type composition, and cell-type specific transcriptomes, especially of neutrophils, should be generated to identify which genes are undergoing changes in expression in individual blood cell types.

It is also important to recognize the limitations of studying a neurodegenerative disease like FRDA by quantifying gene expression in peripheral blood. The fold changes and regression coefficients with FDS we observed are quite small in magnitude when compared with what is typically observed in model systems and post-mortem studies. Due to the scale of the study, we were not able to control some factors associated with the sample collection that could increase the variability in our gene expression signal, such as fasting, exercise and the time of day the sample was collected. We cannot determine conclusively whether these factors may have confounded our study but we anticipate that they would likely reduce our power to detect effects, making our results more conservative.

The inflammation occurring in FRDA is not an acute response to an infection or a traumatic injury; instead it is likely to be similar to the chronic low-grade inflammation observed in other neurodegenerative and inflammatory disorders (17). Our large sample size and rigorous correction for potential confounders have provided the statistical power to identify a broad inflammatory signature. No individual gene can fully quantify the inflammatory response and other cellular pathology, but in aggregate these genes provide insights into the effects of *FXN* deficiency. A further strength of our large sample size is that we can capture more of the genetic variation across FRDA patients than is logistically feasible in model systems and post-mortem studies, which makes our results relevant for a broader range of patients.

Future studies of FRDA in humans should characterize the peripheral inflammatory state of FRDA patients, and seek to identify whether this inflammation contributes to the pathology of the disease, or is merely a response to stresses induced by it. In particular, proteomic cytokine profiling and immune cell activity assays could provide valuable biomarkers beyond gene expression.

Materials and Methods

The full pipeline and code used for all of the analyses is available on Github (https://github.com/coppolalab/FRDA_pipeline) and a summary is provided in this section.

RNA collection and microarray hybridization

Peripheral blood was collected in Paxgene tubes and frozen before RNA extraction, which was performed using a semi-automated system (Qiacube). Subjects were not specifically instructed to fast or refrain from exercise, and the time of collection was not uniform. RNA quantity was assessed with Nanodrop (Nanodrop Technologies) and quality with the Agilent Bioanalyzer (Agilent Technologies), which generated an RNA integrity number (RIN) for each sample. Total RNA (200 ng) was amplified, biotinylated and hybridized on Illumina HT12 v4 microarrays, as per manufacturer's protocol, at the UCLA Neuroscience Genomics Core. Slides were scanned using an Illumina BeadStation and signal extracted using the Illumina BeadStudio software (Illumina, San Diego, CA).

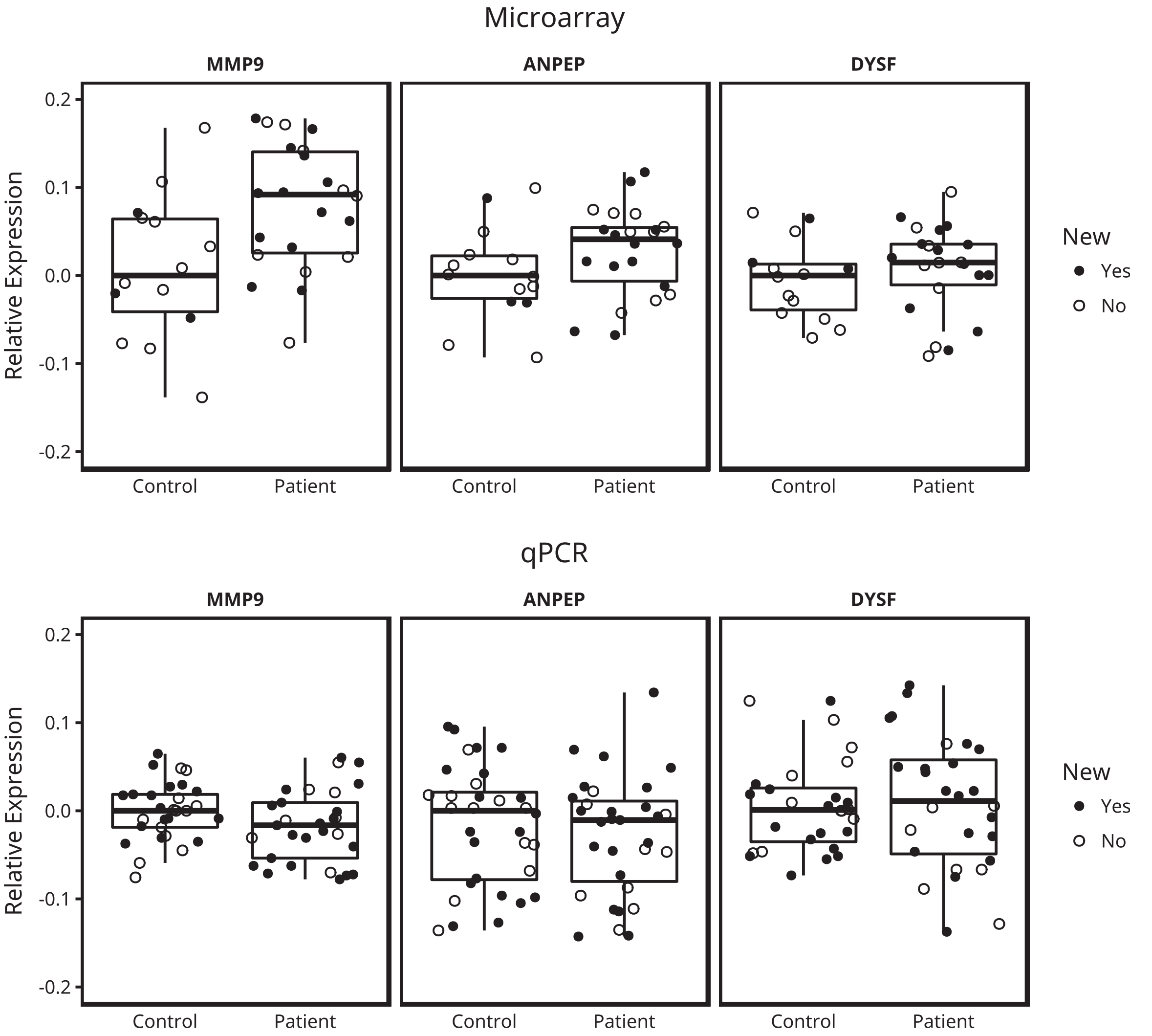


Figure 9. qPCR and array validation of top three DE genes in 32 patients and 32 age- and sex-matched controls. Boxplots showing the relative expression of the top three DE genes to the median value of the control samples. A total of 21 patient and 16 control samples (marked with closed circles) were new and not previously included in the analysis. Top: microarray data, bottom: qPCR.

Array preprocessing

Array preprocessing was performed using the standard pipeline from the *lumi* package (22), which is designed specifically for Illumina microarrays. Raw intensities were normalized using variance-stabilized transformation (23) and interarray normalization was performed with robust spline normalization. A total of 17 outliers were removed from the full dataset and 6 outliers were removed the patient-only dataset using sample-wise connectivity z-scores. Batch effect correction was performed using ComBat from the *sva* package (24). Probes were filtered by detection score and unannotated probes were dropped. Duplicate probes for the same gene were dropped using the maxMean method with the *collapseRows* function (25) from the *WGCNA* package, which only keeps probes with the highest mean expression across all of the samples. After all probe filtering steps, 16 099 probes were used for analysis of the full dataset, and 15 198 probes for the patient-only dataset.

Removal of confounding covariates

Age and sex were found to be collinear with disease status (Supplementary Material, Fig. S1). To account for this, the effects of both covariates were fitted and removed using the median posterior estimates from linear models for each gene made with the *BayesFactor* package (26,27).

Differential expression

Differential expression between patients, carriers and controls was assessed using Bayesian model comparison on linear models for each gene generated with the *BayesFactor* package (26,27). Bayesian model comparison produces Bayes factors (BFs) instead of *P*-values for assessing significance. A BF is the ratio of the probabilities of two models, and reflects the amount of information gained in terms of variance explained when adding one or more variables to a model. Because age and sex were already removed due to collinearity, only disease status and RIN were available to use as variables. The full model containing the intercept, disease status and RIN was compared with the null model containing only the intercept and RIN. BFs were log-transformed to log BFs to place them on a more practical scale (28), and a log BF of 0.5 was used as a cut-off for significance of the alternative model to the null model (29). Although we are fitting a separate model for each gene and thus running thousands of tests, BFs do not require adjustment for multiple comparisons because they are model comparisons (26).

Posterior estimates of the regression coefficients were generated using 10 000 iterations of Monte Carlo Markov chain sampling with a random seed set to 12 345 to guarantee reproducibility. We then specified three contrasts: patient-control, patient-carrier and carrier-control. For contrasts, the posterior samples were subtracted from each other in the order specified to produce an estimate of the difference in expression between the two groups. The median of this estimate was treated as the log fold change (log FC). The pp of the pairwise comparison being in the same direction as the log FC was defined as the number of posterior samples that were non-zero and had the same sign as the log FC.

The Bayesian FDR for each pairwise comparison is $1 - \text{pp}$ of the comparison, so we used a pp of 0.95 as our threshold for pairwise significance, so that the FDR for individual genes would be less than 5%. The global FDR for a pairwise comparison was computed by taking the mean of the FDR values for all of the genes that were found to be significantly DE for that comparison (adapted from 30, 31).

Regression with functional disability stage and other phenotypic measures

Several phenotypic measures were available in a large subset of the FRDA patients ($n = 308$), including FDS from the FARS, the shorter of the two GAA repeat expansions (GAA1), and the disease duration (the difference between age of onset and age at draw). Patients that were compound heterozygotes, with one loss-of-function *FXN* variant on one allele and a repeat expansion on the other, were excluded from this analysis. Age was found to be collinear with all three measures (Supplementary Material, Fig. S2) and was removed using the same linear modeling with *BayesFactor* described previously.

Similar to the approach used for differential expression, linear models for each gene were fitted using *BayesFactor*. The full model containing the intercept, the continuous phenotype (FDS, GAA1 or disease duration), sex and RIN, was compared with the null model without the continuous phenotypes and log BFs were computed. Posterior estimates of the coefficients were generated using the same parameters described above, and posterior probabilities were defined as the number of samples in an estimate that were non-zero and whose sign was opposite that of the median estimate. The same thresholds of log $\text{BF} > 0.5$ and $\text{pp} > 0.95$ were used to assess significance of the linear relationship between gene expression and the continuous phenotypes, and the global FDR was computed as described for differential expression.

Gene coexpression network analysis

Weighted gene co-expression network analysis (WGCNA) was run on (1) the full set of samples; and (2) the subset of patients with complete phenotypic information described above. Only batch effect was removed using ComBat, as the network construction step must be performed on data that has not any source of biological variation removed. The pipeline from the *WGCNA* package was used as reported previously (9). A signed network with a soft power of 6 was generated, and a module dissimilarity threshold of 0.2 was used to merge correlated modules. Hub genes were identified in network modules using scaled connectivity, the ratio of a specific gene's within-module connectivity to the maximum within-module connectivity in that module.

Eigengene values, summarizing gene expression within each module, were compared across disease status using the same linear model approach described for differential expression, with age and sex being regressed out before fitting the final models. Posterior estimates of the model parameters were generated using the same parameters as described previously. Similar to the approach used for genes, module eigengenes with a log $\text{BF} > 0.5$ when comparing the alternative model to the null were considered different across conditions, and pairwise comparisons were also considered significant if their 95% credible intervals did not overlap. For regression with continuous phenotypes, the same linear modeling, removal of age effect and posterior estimation as that described for regression of genes was used with the module eigengenes. An eigengene with log $\text{BF} > 0.5$ was considered to have a significant linear relationship with the continuous phenotype.

Overlap with other datasets

We compared our results with two other human datasets from Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/): GSE11204 (32) and GSE30933 (33), as well as a dataset generated on a novel mouse model of frataxin deficiency (8). The same workflow used to identify DE genes in our data was applied to these datasets, with adjustments made to account for platform differences. Complete descriptions of the datasets and analytic procedures are available in the Supplementary Material. Enrichment was tested using the log BF computed from a hypergeometric overlap test (34) implemented in *BayesFactor*.

Cell type deconvolution

Cell type deconvolution was performed using the quadratic programming method (35) implemented by the *CellMix* package (13), which provides a peripheral blood dataset (14) that can be used to estimate proportions of cell types in transcriptomic data. Deconvolution was run on the raw, unprocessed array data as recommended, although outliers were removed so that only the samples used in the final analysis were used to compute cell type proportions. The proportions were separately estimated in the full group of patients, carriers and controls, as well as the subset of patients used for phenotype regression.

The significance of differences in proportions of cell types across patients, carriers and controls was separately assessed for each cell type using the same Bayesian model comparison and pp estimation described for differential expression. The effects of age and sex were removed by linear regression from the raw expression data before running *CellMix* as described for differential expression, as both variables were confounded with disease status. The significance of regression of FDS with cell type proportion was also determined using the same Bayesian model comparison and pp estimation described for differential expression. The effect of age was removed by linear regression from the raw expression data before running *CellMix* as described for phenotype regression because it was collinear with FDS.

Gene set annotation

Enrichment of genes for specific ontologies and pathways was analyzed using the following datasets downloaded from Enrichr

(36,37) (RRID: SCR_001575): GO Biological Process 2015 (RRID: SCR_002811), GO Molecular Process 2015 (RRID: SCR_002811), KEGG 2016 (RRID: SCR_012773) and Reactome 2016 (RRID: SCR_003485). Enrichment scores were computed using a log BF obtained using the same hypergeometric contingency table implemented in *BayesFactor* (34) used for overlap testing.

qPCR validation

Taqman qPCR was used to validate expression changes observed for the top three genes, in 32 patients and 32 age- and sex-matched controls. 8/32 (25%) patients and 11/32 (34%) controls were new subjects that had not been studied previously, therefore in addition to being a technical validation, this is also partly a biological confirmation of our findings. RNA was converted to cDNA using the Invitrogen Superscript III First-Strand Synthesis System. The TaqMan TM Gene Expression Assay was then used to detect gene expression in the following three target genes: *MMP9* (Taqman, Hs00957562_m1), *ANPEP* (Taqman, Hs00174265_m1) and *DYSF* (Taqman, Hs01002513_m1). *RPLP0* (Taqman, Hs99999902_m1), *GAPDH* (Taqman, Hs02758991_g1) and *β-Actin* (Applied Biosystems, 4326315E) were used as reference genes. Three technical replicates were performed for each reaction, resulting in nine replicates for each biological sample, for a total of 576 PCR amplifications. The real-time PCR was carried out on a LightCycler 480 (Roche) instrument and the C_t values were retrieved using the instrument software.

C_t values for the three targets genes and three reference genes were normalized to a dilution curve as described previously (38) and outliers were identified and removed in two steps. First, data were standardized by subtracting the mean and dividing by the median absolute deviation for each pair target and reference genes separately (i.e. only *MMP9* with *RPLP0* as reference). Any reaction with a standardized score with absolute value great than 2 was excluded, resulting in a total of 44/576 *MMP9* reactions, 43/576 *ANPEP* reactions and 50/475 *DYSF* reactions being excluded. After removing these outliers, the median value across all remaining technical replicates for each gene in each subject was computed. Median expression values per subject were again standardized by median and MAD and any subject whose standardized score had an absolute value greater than 2 was excluded. This resulted in 6 subjects being excluded for *MMP9*, 1 subject for *ANPEP*, and 8 subjects being excluded for *DYSF*. The significance of the difference in expression between patients and controls for each gene was assessed using the Mann–Whitney *U*-test because the expression values were not normally distributed. Data from corresponding arrays was processed using the same array preprocessing pipeline as described previously, except that age and sex were not regressed out because they were no longer confounded with disease status.

To maintain consistency with the qPCR analysis, the Mann–Whitney *U*-test was also used to assess the significance of the differences between patients and controls for each gene in the array data. For 9 patients and 11 controls, the array used was from a different time point than the one analyzed in the original DE analysis, providing both technical and biological validation for those subjects.

Data Availability

All raw gene expression data is available for download in NCBI Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/gds) under accession number GSE102008. An interactive differential expression analysis interface for the data is available in the REPAIR database (https://coppolalab.ucla.edu/account/login). Finally, interactive visualizations of our network analysis are available on our website (https://coppolalab.ucla.edu/gclabapps/nb/browser?id=FRDA_Gene%20Expression%20Network%20-%20Diagnosis;ver=, https://coppolalab.ucla.edu/gclabapps/nb/browser?id=FRDA_Gene%20Expression%20Network%20-%20FDS;ver=).

Ethics Statement

Protocols for acquisition of data from subjects were approved by the Institutional Review Boards of UCLA and CHOP, and consent for data to be used was obtained from subjects or the appropriate legal guardian.

Supplementary Material

Supplementary Material is available at *HMG* online.

Acknowledgements

We thank the UCLA Neuroscience Genomics core (http://www.semel.ucla.edu/ungc) for assistance with gene expression data generation, Mr. Sandeep Deverasetty for assistance with the REPAIR database, and Mr. Yining Zhao with assistance with network data visualization.

Conflict of Interest statement. None declared.

Funding

This work was supported by the Muscular Dystrophy Association (grant MDA186994 to GC), the Friedreich's Ataxia Research Alliance (FARA, FACOMS/Friedreich's Ataxia Natural History Study, and grant 20175148). We acknowledge the support of the NINDS Informatics Center for Neurogenetics and Neurogenomics (P30 NS062691) and the Genetics and Genomics Core of the Semel Institute of Neuroscience at UCLA which is supported by IDDRC grant number U54HD087101-01 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development. Funding to pay the Open Access publication charges for this article was provided by Dr. Coppola's grant support from the Friedreich's Ataxia Research Alliance.

References

1. Cossée, M., Puccio, H., Gansmuller, A., Koutnikova, H., Dierich, A., LeMeur, M. *et al.* (2000) Inactivation of the Friedreich ataxia mouse gene leads to early embryonic lethality without iron accumulation. *Hum. Mol. Genet.*, **9**, 1219–1226.
2. Gottesfeld, J.M., Rusche, J.R. and Pandolfo, M. (2013) Increasing frataxin gene expression with histone deacetylase inhibitors as a therapeutic approach for Friedreich's ataxia. *J. Neurochem.*, **126**, 147–154.
3. Pastore, A. and Puccio, H. (2013) Frataxin: a protein in search for a function. *J. Neurochem.*, **126**, 43–52.
4. Cnop, M., Mulder, H. and Igoillo-Esteve, M. (2013) Diabetes in Friedreich ataxia. *J. Neurochem.*, **126**, 94–102.
5. Campuzano, V., Montermini, L., Molto, M.D., Pianese, L., Cossee, M., Cavalcanti, F., Monros, E., Rodius, F., Duclos, F., Monticelli, A. *et al.* (1996) Triplet repeat expansion Friedreich's ataxia: autosomal recessive disease caused by

an intronic GAA triplet repeat expansion. *Science*, **271**, 1423–1427.

6. Lazaropoulos, M., Dong, Y., Clark, E., Greeley, N.R., Seyer, L.A., Brigatti, K.W., Christie, C., Perlman, S.L., Wilmot, G.R., Gomez, C.M. *et al*. (2015) Frataxin levels in peripheral tissue in Friedreich ataxia. *Ann. Clin. Transl. Neurol.*, **2**, 831–842.
7. Bürk, K., Schulz, S.R. and Schulz, J.B. (2013) Monitoring progression in Friedreich ataxia (FRDA): the use of clinical scales. *J. Neurochem.*, **126**, 118–124.
8. Chandran, V., Gao, K., Swarup, V., Versano, R., Dong, H., Jordan, M.C. *et al*. (2017) Inducible and reversible phenotypes in a novel mouse model of Friedreich's Ataxia. *eLife Sci.*, **6**: e30054.
9. Langfelder, P., Cantle, J.P., Chatzopoulou, D., Wang, N., Gao, F., Al-Ramahi, I., Lu, X.-H., Ramos, E.M., El-Zein, K., Zhao, Y. *et al*. (2016) Integrated genomics and proteomics define huntingtin CAG length–dependent networks in mice. *Nat. Neurosci.*, **19**, 623–633.
10. Seyfried, N.T., Dammer, E.B., Swarup, V., Nandakumar, D., Duong, D.M., Yin, L., Deng, Q., Nguyen, T., Hales, C.M., Wingo, T. *et al*. (2017) A multi-network approach identifies protein-specific co-expression in asymptomatic and symptomatic Alzheimer's disease. *Cell Syst.*, **4**, 60–72.e4.
11. Wu, Y.E., Parikshak, N.N., Belgard, T.G. and Geschwind, D.H. (2016) Genome-wide, integrative analysis implicates microRNA dysregulation in autism spectrum disorder. *Nat. Neurosci.*, **19**, 1463–1476.
12. Parikshak, N.N., Swarup, V., Belgard, T.G., Irimia, M., Ramaswami, G., Gandal, M.J. *et al*. (2016) Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. *Nature*, **540**, 423–427.
13. Gaujoux, R. and Seoighe, C. (2013) CellMix: a comprehensive toolbox for gene expression deconvolution. *Bioinformatics*, **29**, 2211–2212.
14. Abbas, A.R., Wolslegel, K., Seshasayee, D., Modrusan, Z. and Clark, H.F. (2009) Deconvolution of blood microarray data identifies cellular activation patterns in systemic lupus erythematosus. *PLoS One*, **4**, e6098.
15. Caielli, S., Banchereau, J. and Pascual, V. (2012) Neutrophils come of age in chronic inflammation. *Curr. Opin. Immunol.*, **24**, 671–677.
16. Gernez, Y., Tirouvanziam, R. and Chanez, P. (2010) Neutrophils in chronic inflammatory airway diseases: can we target them and how? *Eur. Respir. J.*, **35**, 467–469.
17. Amor, S., Peferoen, L.A.N., Vogel, D.Y.S., Breur, M., van der Valk, P., Baker, D. and van Noort, J.M. (2014) Inflammation in neurodegenerative diseases–an update. *Immunology*, **142**, 151–166.
18. Shenton, D., Smirnova, J.B., Selley, J.N., Carroll, K., Hubbard, S.J., Pavitt, G.D., Ashe, M.P. and Grant, C.M. (2006) Global translational responses to oxidative stress impact upon multiple levels of protein synthesis. *J. Biol. Chem.*, **281**, 29011–29021.
19. Mastrokolias, A., Ariyurek, Y., Goeman, J.J., van Duijn, E., Roos, R.A.C., van der Mast, R.C., van Ommen, GJan. B., den Dunnen, J.T., 't Hoen, P.A.C., van Roon-Mom, W.M.C. *et al*. (2015) Huntington's disease biomarker progression profile identified by transcriptome sequencing in peripheral blood. *Eur. J. Hum. Genet.*, **23**, 1349–1356.
20. Liu, J., Aoki, M., Illa, I., Wu, C., Fardeau, M., Angelini, C., Serrano, C., Urtizberea, J.A., Hentati, F., Hamida, M.B. *et al*. (1998) Dysferlin, a novel skeletal muscle gene, is mutated in Miyoshi myopathy and limb girdle muscular dystrophy. *Nat. Genet.*, **20**, 31–36.
21. Polymeropoulos, M.H., Lavedan, C., Leroy, E., Ide, S.E., Dehejia, A., Dutra, A., Pike, B., Root, H., Rubenstein, J., Boyer, R. *et al*. (1997) Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. *Science*, **276**, 2045–2047.
22. Du, P., Kibbe, W.A. and Lin, S.M. (2008) lumi: a pipeline for processing Illumina microarray. *Bioinformatics*, **24**, 1547–1548.
23. Lin, S.M., Du, P., Huber, W. and Kibbe, W.A. (2008) Model-based variance-stabilizing transformation for Illumina microarray data. *Nucleic Acids Res.*, **36**, e11–e19.
24. Johnson, W.E., Li, C. and Rabinovic, A. (2007) Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*, **8**, 118–127.
25. Miller, J.A., Cai, C., Langfelder, P., Geschwind, D.H., Kurian, S.M., Salomon, D.R. and Horvath, S. (2011) Strategies for aggregating gene expression data: The collapseRows R function. *BMC Bioinformatics*, **12**, 322.
26. Rouder, J.N., Morey, R.D., Speckman, P.L. and Province, J.M. (2012) Default Bayes factors for ANOVA designs. *J. Math. Psychol.*, **56**, 356–374.
27. Rouder, J.N. and Morey, R.D. (2012) Default Bayes factors for model selection in regression. *Multivariate Behav. Res.*, **47**, 877–903.
28. Jeffreys, H. *Theory of Probability*. Oxford University Press, 1961.
29. Kass, R.E. and Raftery, A.E. (1995) Bayes factors. *J. Am. Stat. Assoc.*, **90**, 773–795.
30. Efron, B., Tibshirani, R., Storey, J.D. and Tusher, V. (2001) Empirical Bayes analysis of a microarray experiment. *J. Am. Stat. Assoc.*, **96**, 1151–1160.
31. Efron, B. (2008) Microarrays, empirical Bayes and the two-groups model. *Stat. Sci.*, **23**, 1–22.
32. Haugen, A.C., Di Prospero, N.A., Parker, J.S., Fannin, R.D., Chou, J., Meyer, J.N., Halweg, C., Collins, J.B., Durr, A., Fischbeck, K. *et al*. (2010) Altered gene expression and DNA damage in peripheral blood cells from Friedreich's ataxia patients: cellular model of pathology. *PLoS Genet.*, **6**, e1000812.
33. Coppola, G., Burnett, R., Perlman, S., Versano, R., Gao, F., Plasterer, H., Rai, M., Saccá, F., Filla, A., Lynch, D.R. *et al*. (2011) A gene expression phenotype in lymphocytes from Friedreich ataxia patients. *Ann. Neurol.*, **70**, 790–804.
34. Jamil, T., Ly, A., Morey, R.D., Love, J., Marsman, M. and Wagenmakers, E.-J. (2017) Default "Gunel and Dickey" Bayes factors for contingency tables. *Behav. Res.*, **49**, 638–652.
35. Gong, T., Hartmann, N., Kohane, I.S., Brinkmann, V., Staedtler, F., Letzkus, M., Bongiovanni, S. and Szustakowski, J.D. (2011) Optimal deconvolution of transcriptional profiling data using quadratic programming with application to complex clinical blood samples. *PLoS One*, **6**, e27156.
36. Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R. and Ma'ayan, A. (2013) Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*, **14**, 128.
37. Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A. *et al*. (2016) Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.*, **44**, W90–W97.
38. Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L. *et al*. (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.*, **55**, 611–622.